A HEALTH TECHNOLOGY ASSESSMENT OF LOVAAS AUTISM TREATMENT: THE ROLE OF EVIDENCE IN LEGAL, HEALTH POLICY AND HEALTH CARE CONTEXTS

by

SANDRA LYNN BEARD

B.S.N., The University of British Columbia, 1997
M.S.N., The University of British Columbia, 1999

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies
(Health Care and Epidemiology)

THE UNIVERSITY OF BRITISH COLUMBIA

APRIL 2007

© Sandra Lynn Beard, 2007
ABSTRACT

In 1998 parents of autistic children launched a Charter of Rights and Freedoms challenge against the Province of BC for failing to fund Lovaas Autism Treatment (LAT) (Auton et al.). Although initially successful, in 2004 the Supreme Court of Canada overturned the lower courts' decisions and rejected the parents' claims for public funding. In addition to the Charter issue, these legal proceedings also highlighted the discourse over judicial policy making and the Courts' interpretation of medical evidence – specifically, the effectiveness of LAT. The use of medical evidence in law had been identified as an issue by both the American Institute of Medicine (IOM) and the Agency for Healthcare Research and Quality (AHRQ). This thesis was designed to expand on the IOM/AHRQ's previous work by investigating the conceptualizations and processes used by law, health policy and health care within the context of the Auton legal proceedings in order to gain an understanding of how each domain seeks, understands and applies evidence. This was accomplished in two parts. First, the legal dimension of a comprehensive health technology assessment (HTA) framework utilized a qualitative grounded theory methodology to examine participant interviews and legal documents. This analysis resulted in a conceptual framework of scientific evidence pathways that further defined, contextualized and dimensionalized the phenomena of seeking, understanding, and applying evidence within the three sectors. Grounded theory proved to be an effective approach for exploring the legal context and serves to broaden the scope of evidence HTA researchers can offer. Second, the effectiveness dimension of the HTA framework employed the methods of systematic review and critical appraisal to investigate the current state of knowledge on the effectiveness of LAT. This analysis concluded that there remains a paucity of rigorously designed studies due to ongoing methodological limitations. Overall, the strength of the body of evidence on LAT was poor and did little to advance its effectiveness claims. Together, these two analyses provided an update on the effectiveness of Lovaas Autism Treatment; insight into how the evidence of LAT was handled in the Auton case; and a depiction of how the sectors of law, health policy and health care conceptualize the evidentiary process in general.
# TABLE OF CONTENTS

Abstract ........................................................................................................... ii
Table of Contents ............................................................................................ iii
List of Tables ................................................................................................... vi
List of Figures ................................................................................................... vii
Acknowledgements ......................................................................................... viii
Dedication ....................................................................................................... ix

## CHAPTER 1 Background, Objectives and Rationale

1.1 Autism Spectrum Disorder ......................................................................... 1
1.2 Early Intensive Behavioural Intervention/Applied Behavioural Analysis ..... 3
  1.2.1 Lovaas Autism Treatment ......................................................... 5
1.3 Precedent Lovaas Research ..................................................................... 7
  1.3.1 Lovaas, 1987 ........................................................................... 7
  1.3.2 McEachin, 1993 ....................................................................... 8
1.4 Critiques of Lovaas & McEachin .............................................................. 11
1.5 BC Autism Intervention Programs ........................................................... 17
1.6 Policy Framework .................................................................................... 20
1.7 Pertinent Legislative Frameworks ............................................................. 24
1.8 Legal Context .......................................................................................... 27
  1.8.1 Auton ......................................................................................... 27
  1.8.2 Anderson .................................................................................... 31
1.9 Canadian Legal System ........................................................................... 37
  1.9.1 Canadian Charter of Rights and Freedoms .................................. 39
1.10 Other Health Care Charter Cases ............................................................ 41
1.11 Legal Reviews of Auton ......................................................................... 48
1.12 Objectives .............................................................................................. 52
1.13 Rationale ............................................................................................... 52
1.14 Research Questions ................................................................................. 53

## CHAPTER 2 Conceptual Framework

2.1 Understanding the Role of Evidence in Law, Health Policy and Health Care ...................................................................................................... 55
  2.1.1 Concept of Evidence ..................................................................... 55
  2.1.2 Opposing Conceptualizations of Evidence .................................... 57
  2.1.3 Conceptual Background ............................................................... 57

2.2 Processes for Seeking, Understanding and Applying Evidence: Legal Context .................................................................................................. 63
  2.2.1 Judicial Policy Making ................................................................. 63
  2.2.2 Legal Discovery ............................................................................ 64
  2.2.3 Rules of Evidence .......................................................................... 65
  2.2.4 Burden of Proof ............................................................................ 68
  2.2.5 Standard of Care: Expert Witnesses & Clinical Practice Guidelines .............................................................................................................. 69

2.3 Processes for Seeking, Understanding and Applying Evidence: Health Policy Context ......................................................................................... 72
  2.3.1 Policy Process: The Policy Cycle .................................................. 72
# 2.3.2 Evidence-based Decision Making: Health Technology Assessment

## 2.4 Processes for Seeking, Understanding and Applying Evidence: Health Care Context

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.1 Peer Review Process</td>
<td>86</td>
</tr>
<tr>
<td>2.4.2 Consensus Development</td>
<td>88</td>
</tr>
<tr>
<td>2.4.3 Evidence-based Medicine</td>
<td>90</td>
</tr>
<tr>
<td>2.4.3.1 Randomized Controlled Trials</td>
<td>98</td>
</tr>
<tr>
<td>2.4.3.2 Alternate Study Designs</td>
<td>99</td>
</tr>
<tr>
<td>2.4.3.3 Clinical Practice Guidelines</td>
<td>99</td>
</tr>
</tbody>
</table>

## 2.5 Conceptual Framework

## CHAPTER 3 Research Design & Methodologies

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Grounded Theory: The Legal Context</td>
<td>104</td>
</tr>
<tr>
<td>3.1.1 Grounded Theory Methodology</td>
<td>104</td>
</tr>
<tr>
<td>3.1.2 QSR N6 qualitative research software</td>
<td>113</td>
</tr>
<tr>
<td>3.2 Systematic Review and Critical Appraisal: The Effectiveness Context</td>
<td>115</td>
</tr>
<tr>
<td>3.2.1 Systematic Review</td>
<td>115</td>
</tr>
<tr>
<td>3.2.2 Study Quality Assessment</td>
<td>118</td>
</tr>
<tr>
<td>3.2.3 Critical Appraisal</td>
<td>122</td>
</tr>
<tr>
<td>3.3.4 Summary of Research Design</td>
<td>124</td>
</tr>
</tbody>
</table>

## CHAPTER 4 Results

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Analysis of the Role of Evidence</td>
<td>126</td>
</tr>
<tr>
<td>4.1.1 Descriptive Analysis of Conceptual Categories</td>
<td>127</td>
</tr>
<tr>
<td>4.1.1.1 Perspective of Evidence</td>
<td>127</td>
</tr>
<tr>
<td>4.1.1.2 Rules of Evidence</td>
<td>128</td>
</tr>
<tr>
<td>4.1.1.3 Process of Change</td>
<td>130</td>
</tr>
<tr>
<td>4.1.1.4 Adjudicators of Evidence</td>
<td>132</td>
</tr>
<tr>
<td>4.1.1.5 Decision-Makers of Evidence</td>
<td>135</td>
</tr>
<tr>
<td>4.1.1.6 Focus of Application of Evidence</td>
<td>139</td>
</tr>
<tr>
<td>4.1.1.7 Judicial Policy Making</td>
<td>140</td>
</tr>
<tr>
<td>4.1.2 Scientific Evidence Pathways</td>
<td>141</td>
</tr>
<tr>
<td>4.2 Analysis of Lovaas Effectiveness</td>
<td>142</td>
</tr>
<tr>
<td>4.2.1 Search Results and Study Selection</td>
<td>142</td>
</tr>
<tr>
<td>4.2.2 Study Quality Assessment</td>
<td>144</td>
</tr>
<tr>
<td>4.2.3 Critical Appraisal of Study Methods</td>
<td>146</td>
</tr>
</tbody>
</table>

## CHAPTER 5 Discussion and Conclusions

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Discussion</td>
<td>164</td>
</tr>
<tr>
<td>5.1.1 The Role of Evidence in Legal, Health Policy and Health Care Contexts: Seeking, Understanding and Applying Evidence</td>
<td>164</td>
</tr>
<tr>
<td>5.1.2 Application of Concepts to a Judicial Policy Making Case Study: Auton et al.</td>
<td>171</td>
</tr>
<tr>
<td>5.1.3 Effectiveness of Lovaas Autism Treatment</td>
<td>174</td>
</tr>
<tr>
<td>5.1.3.1 Descriptive Data Synthesis</td>
<td>174</td>
</tr>
<tr>
<td>5.1.3.2 Clinical Heterogeneity</td>
<td>174</td>
</tr>
<tr>
<td>5.1.3.3 Methodological Heterogeneity</td>
<td>176</td>
</tr>
<tr>
<td>5.1.3.4 Heterogeneity in Results</td>
<td>178</td>
</tr>
<tr>
<td>5.1.3.5 The Current State of Knowledge on Lovaas Effectiveness Evidence</td>
<td>178</td>
</tr>
<tr>
<td>5.1.3.5.1 Advances and Improvements</td>
<td>179</td>
</tr>
</tbody>
</table>
5.1.3.5.2 Ongoing Methodological Issues .............................................................................. 179
5.1.3.5.3 Included Studies' Reported Limitations ............................................................... 180
5.1.3.5.4 Unanswered Research Questions ....................................................................... 181
5.1.3.6 Strength of Body of Evidence .................................................................................. 182
5.1.3.6.1 Quality of the Included Trials ............................................................................ 182
5.1.3.6.2 Size and Significance of Observed Effects ......................................................... 182
5.1.3.6.3 Consistency of Effects Across Trials ................................................................. 184
5.1.3.6.4 Dose-response Relationships ........................................................................... 184
5.1.3.6.5 Other Plausible Competing Explanations ......................................................... 184
5.2 Conclusions .................................................................................................................... 186
5.3 Strengths and Limitations ............................................................................................. 189
5.4 Further Research ........................................................................................................... 189
References ......................................................................................................................... 191
Appendices .......................................................................................................................... 205
Appendix A The Role of Evidence in Legal, Health Policy and Health Care Contexts .......... 205
Appendix B Scientific Evidence Pathways: Initial Framework ........................................... 206
Appendix C Participant Letter of Introduction .................................................................... 207
Appendix D Participant Consent Form ............................................................................... 208
Appendix E Search Strategy ............................................................................................... 210
Appendix F Study Selection Form ..................................................................................... 219
Appendix G Study Quality Assessment Tool for RCTs: Modified Jadad Scale ....................... 221
Appendix H Study Quality Assessment Tool for Observational Studies: Reisch ..................... 223
Appendix I Data Extraction Form ....................................................................................... 228
Appendix J Seeking, Understanding and Applying Evidence: Key Concepts and Findings .......... 237
Appendix K Interview Questions ......................................................................................... 243
Appendix L Scientific Evidence Pathways: Revised Framework ........................................ 249
Appendix M Excluded Retrieved Papers ........................................................................... 250
Appendix N Summary of Included Studies ........................................................................ 254
Appendix O Analysis of Heterogeneity Between Groups .................................................... 272
Appendix P Direction of Biases .......................................................................................... 279
Appendix Q UBC Research Ethics Board Certificate of Approval ....................................... 281
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Summary of legal cases</td>
<td>36</td>
</tr>
<tr>
<td>Table 2.1</td>
<td>Policy cycle</td>
<td>72</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Decision making strategies</td>
<td>78</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Comprehensive health technology assessment framework</td>
<td>83</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Study selection criteria</td>
<td>116</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>AHRQ coding criteria</td>
<td>118</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Reisch quality assessment tool: revised criteria</td>
<td>122</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Summary of reasons for exclusion of studies</td>
<td>143</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Reisch quality assessment indices for the included observational studies</td>
<td>144</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Jadad quality assessment indices for the included randomized controlled trials</td>
<td>145</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Statistically significant mean between-group differences at follow-up</td>
<td>183</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>The context, actors and level of application of the evidentiary process</td>
<td>101</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>The indicator-concept model</td>
<td>109</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Study Design: The legal and effectiveness contexts of the comprehensive HTA framework</td>
<td>124</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>The context, actors and focus of application of the evidentiary process (revised)</td>
<td>126</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would be remiss if I did not express my sincere gratitude to the key people who assisted me with this thesis. Many thanks to Mary-Doug Wright for her proficiency in systematic literature searches and Carolyn Green for her assistance as my second reviewer. Special appreciation also goes to my thesis committee - Anthony Sheppard, for guiding me through the unfamiliar maze of law and Ken Bassett, for his prowess with systematic reviews. And most importantly, my sincere thanks to Arminee Kazanjian, my thesis supervisor, for her kindness and patience as she guided me towards my vision for my research.
DEDICATION

It is with great love and admiration that I dedicate my thesis to my husband Jonathan and our children Jordan and Devon. During difficult times, both academically and personally, their unconditional love and support were what kept me focused. I am also extremely grateful to a late friend – Judi Moore – who sparked my interest in further education and continuously challenged me to reach for higher goals. And to my parents – Earle and Lilian Fleming – for their genuine interest in my work and their encouragement to never give up on a dream.
CHAPTER 1: BACKGROUND, OBJECTIVES AND RATIONALE

Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a constellation of three pervasive developmental disorders (PDD) of childhood: (a) autistic disorder (AD), (b) Asperger syndrome, and (c) pervasive developmental disorder-not otherwise specified (PDD-NOS). Other conditions that comprise the broader definition of PDD include Rett syndrome and childhood disintegrative disorder (CDD). Autism was first described in 1943 by Dr. Leo Kanner, when a small group of children were found to display abnormal characteristics such as extreme aloofness, indifference to other people, minimal eye contact, severe language deficits, atypical responses to stimuli, minimal pretend or imaginative play, and lack of desire to communicate (American Academy of Pediatrics, 2001). However, it was not until 1980 when the term "infantile autism" was first recognized as a diagnostic criterion in the Diagnostic and Statistical Manual of Mental Disorders (DSM). Since then, the definition and criterion have been expanded to incorporate milder and more common forms of this condition.

PDD are lifelong neurological disabilities of unknown etiology that generally result in variations of impaired socialization, communication, and behaviour. Because autism is a spectrum disorder, symptoms are a heterogeneous mix with distinct variations among children. There is no universal developmental delay or behaviour that classically depicts autism. Very young children typically present with symptoms of delayed socialization such as poor eye contact and a disinterest in their surroundings. They often appear to be inwardly focused and resist physical contact. Communication is delayed and once it appears it is rote, repetitive, and lacking purpose. These children seldom interact with others and display behaviours such as rocking, flapping of their hands and rituals such as repetitively stacking blocks. Intelligence can vary from severe mental retardation to superior intellectual abilities. More specifically, AD is characterized by developmental delays in social interaction, communication, and imaginative play. These children typically demonstrate stereotyped behaviours, interests and activities. Although these behaviours occur in late infancy, there are cases where developmental milestones that were once achieved regress or are lost in their entirety. Studies are currently underway to identify early signs of autism (Zwaigenbaum et al., 2005). Some of this preliminary research indicates that by the age of 12 months infants who were later diagnosed as autistic differed from other infants in the following areas: (a) behavioural markers such as atypical eye contact, orienting to name, social smiling; (b) prolonged latency to disengage visual attention; (c) temperament (passivity followed by extreme distress reactions); and (d) delayed expressive and receptive language.
Asperger syndrome, on the other hand, is limited to impairments in social interactions such as peer relationships, a lack of empathy, and obsessions about certain topics. Language development is not affected and Asperger's children are typically of average to above average intelligence. It is not yet known whether Asperger is actually a higher functioning form of autism or indeed a unique condition of its own. The third ASD – pervasive developmental disorder not otherwise specified (PDD-NOS) – is an atypical form of autism that is often used as a diagnosis when children do not clearly fall within the defined criteria of the other categories. With all three classifications of ASD, a multi-disciplinary team can usually make a diagnosis by the time a child is two years old (Lord, 1995; Lord & McGee, 2001). The *Diagnostic and Statistical Manual of Mental Disorders* (DSM IV-TR) lists twelve criteria for autistic disorder. These criteria are divided into: (a) impairments in qualitative social interactions, (b) impairments in qualitative communication, and (c) restricted, repetitive and stereotyped patterns of behaviour, interests and activities. Because the DSM IV-TR criteria were developed for children 3 years of age and older, attempts to diagnose younger children may be difficult until all characteristics are demonstrated. Therefore, a diagnosis of PDD-NOS may be proffered until the child “grows into” the full diagnosis of AD. However, assigning a preliminary diagnosis remains somewhat controversial. On the one hand, while there is some consensus that the earlier intervention is begun the more beneficial the outcomes, this must be balanced against the social and financial impacts of a false positive diagnosis.

There are conflicting opinions on prevalence and incidence rates for autism. One review of twenty-three English language epidemiological surveys published between 1966 and 1998 found that prevalence rates increased each successive publication year (Fombonne, 1999). In that review, the median prevalence rate for autism rose from 5.2 per 10,000 for studies published before 1989 to 7.2 per 10,000 for studies after 1989, while the incidence of autism remained stable at 0.2% (with rates increasing to 3-7% for siblings of autistic children). However, the author contends that claims of an autism epidemic are unfounded due to methodological limitations of the existing data, changes in case definition, improved case recognition, broader concept of autism, and recognition of autism amongst normally intelligent people (Fombonne, 2001). An update to the 1999 Fombonne review saw the prevalence rate rise to 10.0 per 10,000 (Fombonne, 2003). However, again the author attributes the rise to improved case definition and awareness of autism. In BC, approximately 262 new cases of ASD are predicted each year, yielding a prevalence of 0.65% (Wellington, 2003). Autism targets males more than females with a ratio of 3.8:1, and is not correlated to any specific ethnic group nor to social factors such as a family’s income, lifestyle or educational level.
Early Intensive Behavioural Intervention/Applied Behavioural Analysis

Early Intensive Behavioural Intervention (EIBI) is a generic term referring to comprehensive and intensive forms of behavioural interventions; another synonymous term is Applied Behaviour Analysis (ABA). With research dating back to the 1960s, EIBI/ABA claims to be one of the most efficacious treatment modalities for children with the life-long affliction of autism (Heflin & Simpson, 1998). ABA is the science of human behaviour. It is an objective, systematic process of applying interventions based on the principles of learning theory in an attempt to improve socially significant behaviours (Association for Science in Autism Treatment, 2004). Behaviour modification enlists three general approaches to treatment: (a) operant conditioning, (b) respondent conditioning, and (c) cognitive conditioning (New York State Department of Health, 1999). Most of the current practices focus on operant conditioning to:

1. Increase certain behaviours
2. Teach new skills (e.g., communication, social skills)
3. Maintain behaviours (e.g., teaching self-control)
4. Generalize or to transfer behaviour from one situation to another (e.g., from the resource room to the mainstream classroom)
5. Restrict or narrow conditions under which interfering behaviours occur (e.g., modifying the learning environment)
6. Reduce interfering behaviours (e.g., self-injury)

(Association for Science in Autism Treatment, 2004)

Since autistic children typically exhibit maladaptive behaviours that can be socially inappropriate, self-injurious, or physically and psychologically demanding on their parents, immediate treatment that focuses on changing this behaviour is the priority. Behavioural intervention is usually the treatment of choice for autistic children.

EIBI/ABA therapy is strictly structured and delivered at high intensity in order to reinforce learned behaviours. Interventions are designed to assist children to obtain and/or alter behaviours through a process of reinforcing adaptive responses and suppressing maladaptive behaviours. Two major assumptions underlie the foundation of EIBI/ABA. The first assumption is grounded in theories of neurobiological development, critical and sensitive periods, attachment theory, infant learning paradigms, prevention theory, and intervention research. This is the assumption that EIBI/ABA should be administered during the preschool years when the brain is most amenable to structural and functional change (Bailey, Aytch, Odom, Symons, & Wolery, 1999). The second set of assumptions is derived from theories of family systems, stress and coping, ecological development, parental roles, and cultural influences, and holds as its core principle the concept of family-centered practice (Bailey, et al., 1999; Guralnick, 2001).

EIBI/ABA can be delivered via three different instructional methodologies: (a) discrete trial training (DTT), (b) naturalistic teaching or errorless teaching, and (c) incidental learning. DTT
involves breaking down skills and teaching them one step at a time with the child mastering the simple skills before progressing on to more complex ones. This takes place in the form of a "training trial" format and consists of four components:

1. The teacher or therapist presents a brief, distinctive instruction or question (stimulus).
2. The instruction is followed by a prompt, if the child needs one, to elicit the correct response.
3. The child responds correctly or incorrectly (the response).
4. The teacher or therapist provides an appropriate "consequence".

(DT) is also known as the "ABC model," whereby each trial consists of: (a) an Antecedent (a "directive"); (b) a Behaviour (a "response"); and (c) a Consequence (a "reaction") (Autism Society of America, 1998). DTT is also referred to as the "Clinical/Prescriptive" method and as "Formal Compliance Training".

The discipline of ABA focuses on reliable measurements and objective evaluation of observable behaviour. Outcomes are operationalized in quantifiable ways in order to measure frequencies and durations of specific behaviours. Evaluation of each individual trial is a critical part of therapy. If a response is not elicited within 5 seconds (or the response is incorrect) that trial ends and a new one begins. Children’s response to therapy is monitored by detailed data collection. Once a skill is mastered (for example, a correct response achieved at least 80% of the time, as judged by two therapists), then that particular skill is added to a maintenance schedule so that learned skills are revisited to prevent regression. In addition, these skills are adapted in order to ensure generalization to different people, settings, and materials (Autism Society of America, 1998).

Behavioural programs can take place in a variety of settings. Typically, ABA begins in the home when children are young, as this is a natural learning environment. However, this approach involves a tremendous amount of input and support from the family. Other programs may advocate a school or clinic-based setting as principles of ABA can be more consistently adhered to. In any case, the program invariably transitions to community and educational settings once the children have learnt the basic behavioural and social skills.

Properly designed and executed ABA programs [should] contain many if not all of the components of effective treatment approaches... namely: individualized instruction tailor-made to address the specific needs of the child, behaviourally based methodology, low student-teacher ratio, early treatment, and family involvement. (Autism Society of America, 1998)

Programs are supervised by an ABA consultant who typically receives preparatory training in a university program. It is important to note that there are many different interpretations and contexts of applied behavioural analysis. Many speech and language therapy programs, in
addition to educational interventions, are also grounded in the principles of behavioural therapy. In addition, some autism intervention programs combine EIBI/ABA with other approaches.

**Lovaas Autism Treatment**

While psychologist Dr. O. Ivar Lovaas refined the Lovaas Autism Treatment (LAT) program at the University of California Los Angeles (UCLA) during the 1980s, its origins are rooted in previous behavioural work with institutionalized children dating back to 1963. An important distinction is that LAT is not synonymous with EIBI/ABA; it is a subset based on similar principles. LAT employs the intensive, one-on-one DTT methodological approach with behavioural interventions carried out for 20-40 hours per week by a team of therapists, parents and eventually, teachers. Its intensity makes it a costly intervention with estimates of between $45,000 - $60,000 CAD per child per year (Ludwig & Harstall, 2000; Supreme Court of BC, 2000). Children begin the program as soon as possible after diagnosis (ideally before age 4) and continue for a two to three year period. The program consists of approximately 500 individual tasks to be learnt. The therapist initially focuses on teaching self-help and receptive language skills, nonverbal and verbal imitation, and appropriate play behaviours. Parental involvement is crucial to provide consistency in managing behaviours. Next, children progress towards learning expressive and early abstract language and interactive play with peers eventually leading towards an advanced application, assimilating the skills for academic tasks, socialization, cause-effect relationships, and observational learning. In its original format, inappropriate behaviours were often managed by aversive methods such as shouting “no” or a slap on the thigh, but, these have since been replaced with strategies such as ignoring and implementing a time-out. Of interest, one study examined the effect of removing negative verbal feedback from the Lovaas program (White, 2000). Results actually showed a trend towards negative feedback increasing skill achievement along with a decrease in the number of maladaptive behaviours.

Lovaas attested that only a practitioner directly linked with the original Lovaas program could claim to apply LAT. Lovaas therapists often carry the designation of “Certified Applied Behaviour Analyst” through specialized training conducted at UCLA. These consultants are master’s prepared graduates who have undertaken a 9-month internship consisting of 30 hours per week of training. Lovaas purports that certification is a critical component in order to maintain quality control of this particular approach to ABA.

It is claimed that, when IBI is implemented at an early age, some autistic children will catch up with their normal peers by grade one (Lovaas, 1987); however, to date there is no reliable evidence to substantiate this claim (Basset et al., 2000; ECRI, 2000; Smith, 1999). The central point of controversy surrounding IBI is the efficacy/effectiveness claims of the various types of programs. While there are numerous published articles on the topic, there is a paucity of methodologically valid studies. One technology assessment of comprehensive treatment
programs in the United States located 528 articles of which only 18 met the inclusion criterion of incorporating a control group (ECRI, 2000). After a subsequent review of the internal, external, statistical conclusions and construct validity of these remaining 18 studies, 13 were found to be “fatally flawed”. Three of these final five studies were behavioural programs based on the Lovaas method. Although the Lovaas group showed an improvement in IQ scores and in some functional parameters, because of study design problems, this gain could not be attributed to the method of treatment.

Another systematic review and critical appraisal (Bassett, et al. 2000) was also successful in uncovering 1200 EIBI abstracts. (This search strategy was not limited to the United States and also incorporated a search for fugitive literature). After the application of the National Institute of Health criteria (for evaluation of primary effectiveness data), the same primary Lovaas studies that had been uncovered by the ECRI report remained. Again, it was determined that “while many forms of intensive behavioural therapy clearly benefit children with autism, there is insufficient, scientifically-valid effectiveness evidence to establish a causal relationship between a particular program of intensive, behavioural treatment, and the achievement of ‘normal functioning’ ” (p. ix).

Smith (1999) conducted a systematic review of early intervention therapies between 1980 and 1999 and uncovered nine “behaviour analytic treatment programs”, five of which were Lovaas-based (Anderson et. al., 1987; Birnbrauer & Leach, 1993; Lovaas, 1987; McEachin et al., 1993; Sheinkopf & Siegel, 1998). Only the Lovaas and the follow-up McEachin studies were based on the original UCLA Lovaas project; the others were partial replication studies that incorporated varying intensities of treatment and levels of therapist training. Although Smith was a co-author of the McEachin study, he generally concurred with previous reviewers’ criticisms of the methodological flaws of these studies. When discussing the replication studies he concluded that under some circumstances early intervention might not yield positive results. However, when speaking about the original UCLA studies, he noted that to date this program had the most favourable outcomes and the strongest methodology. Yet, he appealed to the research community to design future studies that were scientifically rigorous.
Precedent Lovaas Research

Lovaas, 1987

In 1987 Lovaas published the results of his study on intensive behaviour modification in the Journal of Consulting and Clinical Psychology. This report (along with its follow-up in 1993 by McEachin et al.) went on to become precedent research studies for the field of applied behavioural analysis. Lovaas' research was the culmination of a behavioural intervention project that began in 1970 at the University of California, Los Angeles. Lovaas' study compared two groups: (a) experimental group (>40 hrs/week of 1:1 intensive treatment); and (b) control #1 (<10 hrs/week of the same 1:1 intensive treatment, plus a variety of other treatments). Control #1 was used to determine the rate of spontaneous improvement in autistic children. Lovaas also incorporated a second control group of children who had been selected from a cohort of children that had previously been studied by Freedman et al. (1985). This group was assembled to counter potential criticisms that children referred to the study might have constituted a subgroup of children with either favourable or unfavourable outcomes. Lovaas described the inclusion criteria for this control group, and stated that these children were treated like Control Group 1 subjects with the exception being that the intervention was administered outside of the Young Autism Project.

Parents of children in both groups received extensive training and were actively involved in their child's treatment plan. Physical aversives were only utilized on the experimental group. Treatment intensity was reduced to 10 hrs/week in kindergarten for those children progressing normally; others continued to receive 40 hrs/week for more than 6 years.

Assignment to treatment groups was not randomized, "due to parent protest and ethical consideration" (p. 4). Instead, assignment was based on the availability of therapists. Lovaas reported, "because fluctuations in staff availability were not associated in any way with client characteristics it was assumed that this assignment would produce unbiased groups" (p. 4). The inclusion criteria consisted of: (a) an independent diagnosis of autism from a medical doctor or PhD psychologist (based on DSM-III criteria), (b) chronological age of less than 40 months if mute and less than 46 months if echolalic, and (c) prorated mental age of 11 months or more at a chronological age of 30 months. Lovaas reported that the latter inclusion criterion resulted in 15% of referrals being excluded. Pre-treatment measures concluded that the experimental group and control #1 were similar on all measures at intake, with the exception of chronological age in which the control subjects were 6 months older on average (this was later shown to be not significant).

Assessments were conducted in a blinded fashion when the children were between 6 and 7 years of age (when they would normally have completed grade one). Outcome measures consisted of IQ scores and first-grade placement data. A child received a score of 3 (and were classified as normal functioning) if they achieved a normal IQ score, had completed
grade one in a normal class in a school for normal children and had been moved forward to
grade two. A score of 2 was assigned to those children who attended grade one in an aphasia
class (language delayed, language handicapped or learning disabled). And finally, a score of 1
was given to children who had been placed in autistic/retarded classes or if the child’s IQ score
fell into the severely retarded range.

Lovaas reported that at follow-up the experimental group was significantly higher than
both control groups on the variables of classroom placement and IQ. Out of the 19
experimental children, 9 (47%) were reported to have successfully completed grade one in a
normal classroom and advanced into grade two. In addition, these children achieved an
average or above average score on IQ tests ($M=107$). Eight children (42%) passed first grade in
an aphasic classroom with IQ scores within the mildly retarded range ($M=70$). And, two children
(10%) attended classes for autistic/retarded children with profoundly retarded IQs <30. There
were no differences reported between the two control groups at either intake or follow-up.
Combined follow-up data from the control groups showed that only 1 child (2%) reached
normal functioning, 18 children (45%) had been taught in aphasia classes, while 21 subjects
(53%) were in classes for autistic/retarded children.

Lovaas conducted a study within a study when he tested the hypothesis that aversives
(shouting “No”, or a slap on the thigh), in response to self-stimulatory or aggressive and
noncompliant behaviour, would produce significant changes. Four children from the
experimental group and another four from Control #1 were initially treated without aversives
and then later, aversives added to the treatment protocol. Behavioural outcomes were
measured and compared at both these time points. Lovaas reported that both groups made
“sudden and stable reduction[s] in inappropriate behaviours and sudden and stable increases
in appropriate behaviours” (p. 7).

Additional findings of the study were that pro-rated mental age was significantly related
to outcome for both the experimental and control #1 groups and that abnormal speech was
significantly related to outcome in control #1. Chronological age at onset of intervention was
not related to outcome (which ruled out the issue of differences in chronological age between
groups at intake). And finally, Lovaas claimed it was possible to predict the nine children that
had achieved normal functioning based on an analysis of eight pre-treatment variables.

**McEachin, Smith & Lovaas, 1993**

A two-part follow-up of the original children from Lovaas’ 1987 study was published in
1993 in the *American Journal on Mental Retardation*. For the first part of the research the
investigators examined whether the gains made by the experimental group had been
maintained after treatment had ended. For the second part, the nine children in the
experimental group that had achieved “best outcomes” were examined for signs of residual
autistic symptomatology (as compared to a cohort of children without behavioural problems – "non-clinical comparison group").

For McEachin et al.'s study, the mean age of the experimental group was now 13 years and the children had been out of treatment for an average of 5 years (range: 0-12 years). More specifically, the group of best outcome children had been without treatment for 5 years on average (range: 3-9 years). In comparison, the mean age of the control group was 10 years with a mean time out of treatment of 3 years (range: 0-9 years). The researchers acknowledged that their experimental group was older and had been out of treatment longer than the control group and explained that the age difference was due to the original treatment assignment procedure. With the Lovaas study, initially all referrals were assigned to the experimental group because therapists were available. However, the authors of McEachin et al. determined statistically that there was no association between the order in which the children were referred and intake or outcome IQ.

Measures obtained for this follow-up study included an assessment of school placement, IQ, the Vineland Adaptive Behaviour Scales, and the Personality Inventory for Children. Then, to compare the best outcome subjects with the non-clinical comparison group, a comprehensive clinical rating scale was developed that captured the typical areas of problems for autistic children. All these tests were administered and scored by blinded assessors for the experimental group, while staff members of the program and outside agencies tested and scored the control group. Statistical tests indicated that there was no significant differences between children who had been evaluated by project staff and those evaluated by outside agencies.

The authors reported that at follow-up, one of the nine best outcome children in the experimental group had regressed and subsequently placed in a special education class. However, one of the other ten experimental children (who had not reached best outcome) had in fact improved and was now in a regular class. The remaining experimental children had maintained their school placement. Therefore, the figure of 47% of children in the experimental group attaining normal classroom placement was upheld. In comparison, none of the nineteen children in the control group were in regular classes. In regard to IQ, the experimental group had a significantly higher mean IQ than did the control group (30 points higher) prompting the investigators to conclude that the experimental group had maintained its gains in intellectual functioning. In addition, the experimental group showed a higher level of functioning, more adaptive behaviours, and less maladaptive behaviours than the control group. And finally, related to personality functioning, the two groups did not differ on overall scale elevation, however, there were differences between groups on two specific scales - the psychosis and the somatic concerns scales – with the control group displaying higher levels of psychosis and lower levels of somatic complaints.
Upon contrasting the best outcome children against the non-clinical comparison group, the best outcome subjects had IQs in the high end of the normal range. In addition, most of these children also scored above average on the Vineland Adaptive Behaviour Scale. However, although there were three best outcomes children that had achieved only marginal scores on some of the Vineland subscales, on average, the group achieved composite scores within the normal range. On the Vineland Maladaptive Behaviour Scale, three children displayed clinically significant maladaptive behaviour, but as a group, the best outcome children did not display clinically significant aberrant behaviour. And finally, in regard to personality functioning, some deviations from average were noted on the personality test and the clinical ratings scales for the best outcomes children. However, the authors contributed this finding to the extreme scores of one child as opposed to problems with the entire group. Although they acknowledged that group averages were not usually interpreted in this manner, they backed up their position by stating that statisticians concur that, “there are many times when group averages represent the performance of few or no subjects within the group” (p. 370). In this situation, they attributed this variance almost exclusively to one subject.

In sum, eight children (42% of the experimental group) “may be judged to have made major and enduring gains and may be described as ‘normal-functioning’” (p. 368) as opposed to none of the control group achieving this goal.
Critiques of Lovaas & McEachin

Lovaas' original 1987 research study, along with McEachin et al.'s (1993) follow-up, continue to be cited and critiqued within academic, legal, and health care forums. Although each respective discipline possesses their own unique criticisms, this next section will discuss some of the methodological issues within an academic context, as this should be the initial step before any health or legal policy decision is considered. Rebuttals by Lovaas and his proponents are also incorporated.

Lack of Randomization

The most often cited criticism of Lovaas' 1987 study was the lack of randomization (Gernsbacher, 2003; Mesibov, 1998; Schopler, Short & Mesibov, 1989). Lovaas' subjects were assigned to either the treatment arm or one of two control groups based on the availability of Lovaas-trained therapists. This was done due to "parent protest and ethical considerations" (p. 4). Therefore, Lovaas developed a "wait-list" control group, which was based on the first-come-first-served principle. If staff members were available when the child was brought in, the child went into the experimental group; if not, the child was assigned to a control group. However, this was not enforced. Patients were not assigned sequentially; it was dependant upon the therapist determining whether they were able to take another child. At that point, the families had an option not to participate.

Lack of randomization brought about concerns regarding the representativeness and comparability of the sample. For example, arguments were made that there were fewer higher functioning children in the control groups. However, Eikeseth (2001), a researcher from one of the Lovaas replication sites in Norway, disputed this by claiming that the experimental group and control group #1 were similar in regard to 19 out of 20 important variables before the treatment was applied. And, as brought up in the Autism legal proceedings, a school of thought exists that while quasi-random assignment could make the treatment administered to the experimental group look better than it actually was, it could not make an ineffective treatment look effective (Baer, 1993).

Inclusion Criteria

Prorated Mental Age

Schopler, Short and Mesibov (1989) took issue with the use of the prorated mental age (PMA) as a selection criterion in determining the intellectual functioning of the children. (Lovaas utilized the Bayley, Binet, and Cattell IQ tests on 90% of his research subjects for pre-testing). "To adjust for variation in MA [mental age] scores as a function of the subject's CA [chronological age] at the time of test administration, PMA scores were calculated for a CA at 30 months (MA/CA x 30)" (Lovaas, 1987, p. 3). Schopler argued that a PMA typically depicts children as lower functioning than the ratio IQ and by using this criterion some children, besides the profoundly retarded, were unnecessarily excluded from Lovaas' study (15% to be exact).
Schopler pointed out that Lovaas' reported mean PMA of 18.8 for his treatment group translated into a ratio IQ of 63, which was significantly higher than any previously reported random sample of autistic children. Jordan, Jones, and Murray (1998) described the use of the PMA as "psychometrically dubious" (p. 111). However, this claim was once again disputed by Eikeseth (2001) by explaining that Lovaas used ratio scores at intake and deviation scores at follow-up. This was a conservative measure as it apparently compared the highest scores at intake with the lowest scores at follow-up.

**Chronological Age**

Lovaas included children <40 months if mute and less than 46 months if they demonstrated the autistic trait of echolalia. Echolalia is recognized as a characteristic of children with a better prognosis (Schopler, et al., 1989). Lovaas acknowledged that the echolalia criterion was employed to select a sample that would progress more quickly, but asserted that data analysis showed no greater improvement for echolalic children.

**Sex Distribution**

Due to the lack of randomization, more girls ended up being enrolled into one of the Lovaas control groups. This claim of bias was based on the fact that females are known to have poorer prognoses than males. Eikeseth (2001) agreed that the experimental group might have had a better prognosis than the two control groups, however, Smith (1997) pointed out that two of the three girls in the experimental group actually achieved the status of best outcome, thus demonstrating that the gender bias criticism was not relevant. However, unequal sex distribution can occur even with randomization, if the study is not large enough.

**Delay in Assessment**

In McEachin's 1993 follow-up study, the experimental group was 3-5 years older than the control group at final assessment. Therefore, gains may have been due to additional development and education. Also, the treatment group had been out of therapy longer. However, Eikeseth (2001) argued this point by claiming that there was no evidence that maturation or non-behavioural education produces large gains in individuals with autism. He went on to suggest that relapse might occur when behavioural treatment is withdrawn. The length of time out of treatment could in fact bias the study against the treatment group.

**Measurements**

**Use of Different Measures Before and After Treatment**

Researchers have also pointed out that Lovaas employed different measures before and after treatment (Howlin, 1997). However, Smith and Lovaas (1997) rebutted this assertion with the argument that this was acceptable practice since no single measure of intellectual functioning could be valid and reliable across an entire range of ages. In fact, Lovaas (2000) responded "few would suggest that 3 year olds should receive the same assessment as 13 year olds" (p. 32). Adjustments for the baseline variable can be made utilizing analysis of covariance.
Measures Not Reflective of Important Areas of Difficulties in Autism

Another criticism made by Howlin (1997) and Jordan et al. (1998) was that Lovaas used gross outcome measures instead of key measures of specific autism features. Specifically, Howlin noted a paucity of measures on social interaction, friendships, conceptual abilities, social communication, obsessional and ritualistic behaviours, and disturbances of mood. Eikeseth (2001) countered this claim by stating that the 1993 follow-up study by McEachin et al. contained 33 outcome measures, all of which reflected important areas of difficulty in autism.

Control Groups

Jordan et al. (1998) contended that Lovaas' control group #2 was an opportunistic sample. However, this was disputed by Eikeseth (2001) by justifying that this group was used to rule out referral and selection bias. Mesibov (1998) argued that different cut-off ages were used for echolalic and mute children; that the control group had fewer higher functioning children than would be expected (typically 20-30% of autistic children are higher functioning); and that each control group received different testing protocols. However, Lovaas dismissed the claim of fewer high functioning children in the control group by arguing that, even with all his experience, he could not predict which child would succeed because high functioning adults may not have appeared as high functioning children (Johnson, 1994).

Treatment

Delay in Treatment

It was noted by Jordan et al. (1998) that control group #1 was delayed in receiving their treatment making the experimental group 6 months older than the control group. (This was the only variable out of 20 assessed that was not consistent between the groups). Eikeseth (2001) agreed with these findings, but argued that Lovaas did not find a relationship between age at intake and treatment outcome when the best outcome children were compared to those children in the study who did not achieve normal intellectual and academic functioning. Due to the young sample (M=35 months), age at intake may not have been an influential variable. However, it does not negate the fact that it may still be an important variable for future studies.

Timing of Initiation of Treatment

While intuitively it makes sense that the earlier treatment is initiated the better the prognosis for an autistic child, Howlin (in 2003) pointed out that in fact there was no research that substantiated this assumption. Lovaas advocated treatment before the age of four, while the Autism Society of America (1998) states, "Although not confirmed by research, it is believed [italics added] that the best age to begin intensive ABA therapy is between 24-42 months or before 3 ½ year of age" (p. 4). While research has demonstrated improvements in children when intervention programs are initiated before the age of four, they had failed to include a systematic comparison with children of other ages. Conversely, there was no evidence to support the hypothesis that delayed initiation of treatment would be of little benefit. Instead, the
belief was “better late than never” as opposed to “early intervention or nothing” (p. 256). Since this time, additional research has been conducted which reported gains for 4-7 year olds after one year of Lovaas treatment (Eikeseth et al., 2002).

**Treatment Fidelity/Integrity**

Ensuring that a treatment is implemented as it was originally conceptualized is an issue that is not only relevant to Lovaas therapy but has plagued researchers in the fields of psychiatry, psychology and education for years. Measurements of treatment integrity were not common when Lovaas conducted his study. However, claims against the integrity of Lovaas’ program (Jordan, et al., 1998) were disputed by Eikeseth (2001) who argued that, although formal treatment integrity was not monitored, specific measures were taken. These included the establishment of treatment protocols and manuals, weekly clinic meetings (with teaching demonstrations), monthly grand rounds, and use of individual logbooks, data sheets and written reports.

**Treatment Intensity**

Jordan et al. (1998) suggested that the outcomes of Lovaas’ study might be due to the intensity of the program as opposed to the actual treatment provided. However, this was later disputed by Eikeseth, Smith, Jahr and Eldevik (2002) who compared the effects of 28 hours per week of behavioural treatment to 29 hours per week of eclectic special education treatment for 4-7 year olds. As previously discussed, the behaviourally treated children appeared to make more progress on measures of intellectual, language and adaptive functioning after one year of treatment.

Lovaas advocated the importance of 40 hours per week of therapy based on the fact that “there are no data to support that an intervention of less than 40 hours per week will result in 47% rate of normal functioning” (Lovaas, 2000, p. 30). However, at a conference of the Autism Society of Connecticut in 1998, Dr. John McEachin recognized the barriers faced by many parents trying to access LAT. His position was that there was no “all or nothing” about the program and if parents could obtain 25 hours per week that would be sufficient (Hultgren, 1998).

**Use of Aversion Therapy**

One of the most contentious issues surrounding Lovaas’ earlier work was the use of aversives. Since these methods are no longer used, one argument is that it invalidates the results of Lovaas’ 1987 study (Jordan et al., 1998). Eikeseth (2001) disagreed and concluded that while it would make it very difficult to replicate and compare studies, the results remain valid.

**Outcomes**

**Claims of Recovery or Cure**

Much controversy has also centered on Lovaas’ claim of achieving “normal intellectual and educational functioning” in 47% of the experimental group (Lovaas, 1987, p. 7). Although Lovaas was cognizant of the debate about whether or not normal functioning equated to
recovery from autism, he still refers to his best outcome subjects as recovered. Somewhere along the line, however, recovered has been translated into cured. Lovaas has publicly rejected the claims of a "cure" by stating "I don't claim a cure because we haven't gotten to the organic variable that is causing autism" (Johnson, 1994, p. 9).

Schopler, Short and Mesibov (1989) joined this argument when they pointed out that the Lovaas study failed to use typical social, behavioural and communication outcome measures to support the hypothesis of attaining normal functioning. They contended that when IQ measures and progress in school are used alone, they are poor indicators of treatment efficacy. School placements may have more to do with administrative policies surrounding special-needs children. In addition, improvements in IQ scores may be due to improved compliance rather than in cognitive functioning. They took issue with Lovaas' statement that "the recovered children show no permanent . . . behavioural deficits and their language appears to be normal" (Lovaas, 1987, p. 8) as the claim was unsupported by data. Mesibov (1998) also pointed out that there are many high-functioning autistic people with near-normal IQ's attending regular public schools who would still be considered severely handicapped.

**Settings**

Lovaas autism treatment, as originally conceptualized, was delivered in a structured clinical setting with daily therapist supervision. Since then, increased demand for LAT has forced the Lovaas Institute to design workshop-based services that can be transferred into community and home settings. Lovaas warned, however, that these workshop-based treatment programs had very different outcomes, and estimated that data documenting normal functioning in these settings would only be about 20 percent (Lovaas, 2000). Lovaas suggested this was due to factors such as high staff turnover, less frequent supervision, lower treatment intensity, and utilizing therapists who had less academic preparation in learning-based theory and research. An even greater concern was that, should these workshop-based services be provided by therapists not trained by UCLA, treatment efficacy would drop to less than 10%.

**Therapists**

Parents seeking behavioural therapists to administer Lovaas therapy are vulnerable to the various claims of expertise made by these practitioners. Lovaas was adamant that only UCLA-trained therapists were qualified to deliver Lovaas autism treatment (Lovaas, 2000). Lovaas therapists had to have a master's degree before they were considered for a 9-month internship that consisted of 30 hours per week of training. However, critics felt that others could be trained in the techniques of discrete trial training (DTT) (Donnelly, 1996).

Of interest is that Lovaas' perspective on behavioural training appears to have changed over the years. In an interview published in The Advocate, Lovaas espoused the virtues of being able to hire and train volunteers stating that, "many people can do this without years of expensive training" (Johnson, 1994, p. 20). In fact, he promoted the two-day UCLA workshop (at
a cost of $1,400) followed by a second workshop two months later (for $600 or $700). Yet, in 2000 while refuting some of the misunderstandings surrounding the UCLA Young Autism Project, Lovaas stated, "reading the teaching manual, attending a workshop led by UCLA certified consultants, practicing behavioural therapy on several families or spending a short time at the UCLA affiliated site, does not make a person qualified to provide UCLA based treatment" (Lovaas, 2000, p. 30).

**Costs**

With the lack of empirical guidance on the most effective intensity of treatment, LAT expenses can fluctuate dramatically. Lovaas estimated the cost of treatment at $60,000 USD per year, or $120,000 for the recommended two years of treatment to achieve "normal functioning", at which point the best-outcome children would not require any additional treatment. He contrasted this against an estimate of more than two million dollars for life-long protective care (Lovaas, 2000).

**Appropriateness For All Autistic Children**

Lovaas determined that his therapy, which is language-based, does not work on children who are visual learners. In fact, his research focus has shifted towards this group of children in order to understand how they learn. "Skill in auditory matching" was the only predictor of success at the end of three months (e.g., whether the children can imitate sounds like words) (Johnson, 1994, p. 19). Those that appeared to benefit the most from LAT were nonverbal and non-compliant (Donnelly, 1996, p. 6).
Historically, the government of British Columbia used to offer a myriad of services for autistic children such as respite, home support services, childcare, physical and occupational therapy, speech/language therapy, hearing services, and behavioural support. However, the focus of these programs was on supportive care as opposed to therapeutic services. Therefore, in 1995, families of autistic children began a letter-writing campaign in an attempt to lobby the government for Lovaas Autism Treatment. By 1996 the parents had collected the names of sixty-three psychiatrists on a petition in support of Lovaas and by 1998 over eight thousand citizens had signed a petition urging the Province to fund effective autism treatment. However, in response, the government placed a moratorium on any further autism therapies; this action precipitated the Auton legal proceedings (to be described in a later section).

In May 1999, one year after the Auton case had commenced, the BC Ministry of Children and Families and the Ministry of Education launched its Autism Action Plan and the Autism Action Implementation Plan to address concerns expressed by the public in regard to autism services for children, youth and their families. This plan was conceptualized after consultation with parents, service providers, community advocates and government staff. The plan was generic in that it provided “early intervention and treatment”; it did not specifically outline a program of early intensive behavioural intervention. However, implementation was hampered by both funding constraints and the equity viewpoint that autistic programs had to be balanced against services to other special needs children.

In March of 2001, in response to the Auton BC Supreme Court ruling (July 2000), an inter-ministerial committee designed the Provincial Centre for Autism and Related Disorders program (P-CARD). The Early Intensive Behavioural Intervention (EIBI) initiative for children under the age of 6 started in May 2001 and entailed contracting three child development centres located in Ladner, the Thompson Okanagan region, and Victoria, to provide the first official EIBI program in the Province. This program offered EIBI for 20 hours per week on a 1:1 basis in addition to speech-language and occupational therapy. Each program had opportunities for the children to integrate with typical peers, involved parents in the training and intervention, and utilized only positive behavioural reinforcement to handle behavioural problems (Mirenda, 2005). In total, 75 children were enrolled in these centres. A component of the program was to include training of additional EIBI therapists in addition to an evaluation project to measure the impact of the therapy on the children, their families and the community.

In May 2002, instead of expanding the EIBI program, the Ministry of Children and Family Development (MCFD) initiated an interim early intensive intervention (IEII) funding option for autism treatment (the program name has subsequently changed to: Autism Funding: Under Age 6). Under this program, families of children with ASD under the age of six are eligible to receive a maximum of $20,000 per year for autism interventions, training, and equipment. However, this is a
parent-administered program that requires families to assume a range of clinical and administrative responsibilities for their children's behavioural treatment programs. Therefore, they must enter into formal funding agreements with the government. Clinical responsibilities include selection of a behavioural therapist, and in consultation, development of a behavioural plan of intervention. In regard to the administrative responsibilities, there are two options for the IEII program. The first is the invoice payment option whereby the family applies to the MCFD for a billing number so that treatments can be invoiced directly to the government. The second option is the direct funding option where the family sets up a trust account for the child and the MCFD deposits funds into the account on a monthly basis. Both these options require detailed record keeping and accounting procedures to verify services received. In addition, the behavioural therapists become employees of the families, thus necessitating full payroll responsibilities such as the requisite submission of employment insurance premiums, taxes, and Work Safe BC contributions.

Although administrating the IEII program can be challenging for some families, the benefit is that the parents have the option of choosing and customizing a program that is specific to their child’s needs. However, to be eligible for funding, the therapists must be selected from a list of provincially approved service providers. This list is maintained by a third party provider – Autism Community Training (ACT) – which also provides information, training and support services to these families.

The autism funding program was further expanded in April 2003 to include school-aged children (initially called the Extended Autism Intervention Funding [EAI] program and then changed to Autism Funding: Ages 6-18). Four Ministries are responsible for collaborating on this program: (a) Health Services/Health Planning (for diagnosis and assessment); (b) Children and Family Development (intervention funding); (c) Education (education programs); and, (d) the overarching Provincial Health Services Authority. For children 6-18 years, diagnosis does not have to be established by a multi-disciplinary team (unlike the under 6 year olds). Again, the family is free to choose the type of behavioural treatment as long as it is administered by an approved service provider and is based on the child's documented intervention plan. However, for these school age children, the funding is only provided for intervention services delivered outside of regular school hours (e.g., the behavioural therapist would not be allowed into the classroom). Currently, the maximum funding available for children 6-18 year old is $6,000 per year. However, recently the BC government announced that it would be increasing its funding to the school districts to $16,000 per year for each student diagnosed with autism spectrum disorder (as opposed to the current funding agreement which only included children diagnosed with autism) (Lee, 2006).

The government also established the BC Autism Assessment Network (BCAAN), under the jurisdiction of the Provincial Health Services Authority (PHSA), which is responsible for the
assessments and diagnosis of all children with suspected autism. The Ministry of Health Planning’s Standards and Guidelines for the Assessment and Diagnosis of Young Children with Autism Spectrum Disorder in British Columbia (BCMOHP, March 2003) provides the criteria for which these children are assessed. In May 2003, the government opened the Provincial Autism Resource Centre (PARC) in Sunny Hill Health Centre for Children in Vancouver followed by centres at the Queen Alexandra Centre for Children’s Health in Victoria and at Children and Women’s Hospitals in Vancouver. These multidisciplinary teams of nurses, psychologists, social workers, pediatricians, psychiatrists, occupational therapists, and speech-language pathologists are responsible for the assessment and diagnosis of autism. In order to receive funding, parents must either use the recommended clinicians from the PARC, or if they choose a private consultant, the clinician must use the same guidelines that PARC follows.

In August 2003, the government’s clinic-managed EIBI program and its parent-managed IIE program were formally evaluated (Mirenda, 2005). In total, 39 children from the EIBI program and 31 from the IIE program were evaluated for up to two years for the IIE program and up to 32 months for the EIBI program. Assessment tools were administered at baseline, 6 months, 1 year, 2 years, and 32 months. The study evaluated the following:

1. Treatment hours and its relationship to progress made by the child.
2. Impact of types of service provided, diagnosis, maternal education, chronological age, and child “testability”.
3. Rate of change prior to and over 2 years of intervention.
4. Parental perceptions and satisfaction.

Study results indicated that there was no significant relationship between a child’s progress and the hours of treatment received; however, some predictors of progress were identified. There was no significant difference in outcomes between the EIBI and the IIE services nor did maternal education or the child’s age at treatment initiation predict the child’s progress. However, children who were “testable” before initiation of treatment, 6 months later on, or both had made more measurable gains than children who were un-testable. For example, the testable children demonstrated an increase in IQ from 60 to 83.7 (un-testable children had a small significant increase from 45.8 to 49.7). In addition, both groups had a significant reduction in autism behaviours. Overall, both programs assisted the children to make more progress per month than they made before they started treatment. And, in regard to parental perceptions and satisfaction, results showed that parental stress was greatly reduced in both types of programs and that they were very satisfied overall.

In order to grasp the multitude of complexities that impact the design and implementation of government programs such as the autism project, it is necessary to step back and examine the underlying policy and legislative frameworks that impact policy development. Therefore, a brief summary of these main concepts follows.
Policy Framework

The field of policy science evolved after World War II when political science students became interested in investigating the relationships between governments and citizens. During this time, students sought frameworks that incorporated the concepts of justice, equity, and the pursuit of social, economic, and political development into their existing constructs of normative or moral dimensions of government. Research of this era focused on the purposes of government and the activities required of governments to ascertain a high quality of life for their people. However, it was generally recognized that a gap in knowledge existed between political theories and actual political practices. This void was filled with the development of new approaches to study political phenomena; hence the evolution of the discipline of policy science (Howlett & Ramesh, 1995).

Harold Lasswell was one of the original pioneers of an approach that went beyond focusing on the structure of governments, the behaviour of political actors, or what governments ought to do. Instead his perspective was directed towards what governments actually do - the public policies and policymaking processes. Lasswell asserted that policy science differed from previous methodologies in that it was multi-disciplinary, problem solving, and normative in its approach. Multi-disciplinary refers to the fact that policy science goes beyond its historically narrow focus on government institutions to include perspectives from fields such as sociology, economics, law, and politics. Problem solving means that research should be relevant with a view to solving actual social problems as opposed to contributing to more academic debates. And, the criterion of normative assumes that, "policy science should not be cloaked in the guise of 'scientific objectivity', but should recognize the impossibility of separating goals and means, or values and techniques, in the study of government actions" (Howlett & Ramesh, 1995, p. 3). Although these three characteristics of policy science have been revised somewhat over the years, essentially Lasswell’s early work remains the framework of modern day studies.

Public policy has many competing definitions, but is essentially "a course of action or inaction chosen by public authorities to address a given problem or interrelated set of problems" (Pal, 1997, p. 1). Its first feature - a course of action - implies that policy-makers develop and work within a policy framework by applying a policy to a specific context. Also of significance to this definition is the assertion that a government's inaction on an issue can also be defined as policy in of itself. However, for inaction to be viewed as a public policy, a government would have had to consciously come to the decision not to act, or to maintain the status quo. The definition’s second feature – referring to problems and interrelated sets of problems – recognizes that public policymaking is a method of dealing with problems; therefore policies are not viewed as ends in themselves but as instruments or tools to address public concerns. "Policies are mental constructs, strings of phrases and ideas. The text of a policy statement and the programs and actions that follow it are simply evidence for the mental construct" (Pal, 1997,
Public policy goes beyond the everyday practice of decision-making; instead policy results from decision-making conducted within a framework or pattern (Pal, 1997). Sometimes this occurs quite insidiously with governments coming to the realization that decisions made over time may in fact now constitute a policy. While individuals, private organizations and governments alike all have policies, a policy is deemed a public policy based on the source of the policy, not on the target of the policy, and can originate from legislature or another delegated authority.

Theorists such as William Jenkins and Thomas Dye debated whether public policy was a process or a choice. Jenkins's conceptualization of public policy was, "a set of interrelated decisions taken by a political actor or group of actors concerning the selection of goals and the means of achieving them within a specified situation where those decisions should, in principle, be within the power of those actors to achieve" (Howlett & Ramesh, 1995, p. 5). This definition acknowledges that rarely does a government solve a problem with one single decision. In reality, a number of decisions would have been made along the way to comprise a public policy. It also recognizes a government's capacity to implement a policy by conceding that policy choices might be limited by factors such as financial constraints or lack of resources. And finally, Jenkins' conceptualization supports the goal-oriented behaviour of policy programs. Here, governments define a goal for a problem and develop the mechanisms in which to achieve it.

Alternatively, Thomas Dye defined public policy as being "anything a government chooses to do or not to do" (Howlett & Ramesh, 1995, p. 4). While somewhat vague, it did stipulate that the agent of the policy must be a government (therefore possessing the legitimate authority to impose policy) and that to be considered public policy it must consist of a deliberate decision. And finally, James Anderson's definition described policy as "a purposive course of action followed by an actor or a set of actors in dealing with a problem or matter of concern" (p. 6). This definition introduced the possibility of multiple sets of policy actors as policies were often formulated not only by multiple decisions, but also by multiple sets of decision-makers. It also recognized the link between government action and the perception of the existence of a problem or issue requiring attention.

There are five categories of policy actors that participate in the policy process: (a) elected officials, (b) appointed officials, (c) interest groups, (d) research organizations, and (e) the mass media. Elected officials include members of the government's executive (also known as the Cabinet) and the legislature. Cabinet ministers play a significant role in policy formation as they are bestowed with the constitutional authority to govern the country/Province. This authority is supported by their legitimate control over information dissemination and financial resources, in addition to their access to the media and bureaucratic expertise. The legislature,
on the other hand, is a forum in which public problems are identified and policies debated. It is a process that holds a government accountable to its people, as opposed to a venue for developing policies. While legislators have the opportunity to question bills and policies, in majority governments the authority of the executive overshadows legislators' input into the actual policymaking process. Hence, legislators are not significant actors in the policy process (Howlett & Ramesh, 1995).

Appointed officials are methodological and substantive experts (known as "bureaucrats") who, in theory, are hired by the executive to assist with government's public policy and administrative duties (not so much in actual practice). This is a heterogeneous, non-partisan collection of multiple disciplines, each with their own unique perspective, values and ideals. The present day reality of complex, multi-sectoral governments has imparted greater responsibilities on the role of the bureaucrat. These appointed officials now assume the policymaking and implementation functions once reserved for the executive and legislature. Since policymaking decisions occur in isolation within these bureaucratic committees, other policy actors are not privy to the process. However, while bureaucracy may appear to reign supreme, it still remains the ultimate responsibility of the executive to approve all policies implemented. Typically, the more contentious the policy, or the more divisive the bureaucracy on the policy's merit, the more executive input is proffered (Howlett & Ramesh, 1995).

Interest groups are outside of the preserve of the government and are typically social organizations that represent groups such as labour forces (e.g., union workers, businesses, professional associations), individuals with special needs (e.g., Families for Effective Autism Treatment), and consumer activists (e.g., BC Taxpayers' Association). Interests groups can play key roles in the policy process, as they are often the purveyors of the information that is required by politicians and bureaucrats for policymaking. Knowing this, interest groups may elect to only release information to the political party that is most sympathetic towards their cause. Alternatively, politicians and opposition parties may seek out interest groups that support their policy platform. Coupled with the promise of information, interest groups also wield power in the form of financial incentives (e.g., campaign contributions) and votes. However, how much power an interest group has will depend upon its organizational structure, for example, its size, financial and human resources, and coalitions with other organizations (Howlett & Ramesh, 1995).

Research organizations are also independent of government and include university researchers and independent "think tanks". University research tends to be non-partisan and specialized, focusing on expanding theoretical frameworks of public policy, while private research organizations have a more comprehensive, applied approach, producing recommendations for policy-makers. Think tanks may be aligned with the ideologies of certain
political parties and will seek them out when disseminating their findings (Howlett & Ramesh, 1995).

And finally, the role of the mass media in the policy process remains somewhat controversial with some people feeling they play a marginal role to others stressing their importance. However, what is agreed upon is that the media links governments to society. Investigative reporters have the mandate of discovering problems, discerning their causes and suggesting remedies. Thus, they often bring problems to the forefront of the political scene. "The role of the media in the policy process lies in the fact that in reporting problems they combine the roles of passive reporter with active analyst as well as an advocate of a solution" (Howlett & Ramesh, 1995, p. 59). Interest groups often take advantage of the media's power by gaining government's attention through sensationalist news reporting. Alternatively, public officials will also seek out the media to release select information that may promote their political platform. Either way, policy actors are wise to the influence of media on popular opinion and will opt to use it when strategically warranted.
Pertinent Legislative Frameworks

The overarching mandate, under which governments must adhere to, is the Canada Health Act. The Hospital Insurance and Diagnostic Services Act of 1957 and the Medical Care Act of 1966 formed the framework for the Canada Health Act, as we know it today. In the late 1950s the federal government offered to compensate the Provinces 50% of their cost for insured health services if they agreed to insure all hospital and physician services and comply with some standard principles. Although these Acts did not prevent the Provinces from extra-billing patients for services, it was not to the Provinces' advantage to do so as their federal contribution would be reduced proportionately. In 1977 a model of block funding that was no longer linked to direct provincial expenditures for health care replaced this formula for shared costs. In response, many Provinces implemented direct patient charges, which the federal government viewed as a threat to the free and universal philosophy of health care. The Canada Health Act was enacted on April 1st 1984 thereby merging the two previous Federal Acts. In addition, it added additional restrictions and penalties against extra-billing of patients.

The primary objective of the Canada Health Act is "to protect, promote and restore the physical and mental well-being of residents of Canada and to facilitate reasonable access to health services without financial or other barriers" (Baker and Bhabha, 2004, p. 25). The Provinces are required to cover "insured health services", defined as all "medically necessary" hospital services, "medically required" physician services, and surgical-dental procedures performed within a hospital. In order to be eligible for the cash transfers and tax point transfers the Provinces had to comply with the five principles of the Act: (a) universality, (b) comprehensiveness, (c) portability, (d) public administration, and (e) accessibility (Flood & Choudhry, 2002; Madore, 2004). Of specific relevance to the Auton case were the principles of universality and comprehensiveness. Universality is a fundamental Canadian value that ensures that qualified provincial residents receive insured health care services provided on uniform terms and conditions. Qualified, as defined under the Medicare Protection Act, designates that a beneficiary of provincial health care services must be a Canadian Citizen (or have permanent residency status), and must reside within BC for a minimum of 6 months per calendar year. Comprehensiveness refers to the provision of all insured health services provided by hospitals, medical practitioners (or dentists within hospitals); or, if the provincial law permits, additional services delivered by other listed health care practitioners.

A crucial issue that unfolded in Auton pertained to the fact that the CHA does not specifically define "medically necessary" or "medically required"; nor are these terms operationally defined in provincial legislation. Typically, which services are included in publicly funded plans are negotiated between the government and medical associations (this is also the case when services get de-listed). Critics argue that this process is flawed in that the process for determining medically necessary services is too intimately linked to determining compensation.
rates for physicians. An example of the limitations of comprehensiveness is in regard to prescription drugs. Although medication is medically necessary, prescription drugs in Canada are only insured at a rate of 36% compared to 91.1% for hospital services and 99% for physician services. This places Canada alongside the United States, Mexico, and Turkey as far as providing a publicly funded universal drug program (Flood & Choundry, 2002).

Also of relevance to Auton is the CHA’s definition of “health care practitioner” as, “a person lawfully entitled under the law of a Province to provide health services in the place in which the services are provided by that person” (Supreme Court of Canada, 2004, Appendix A). This was at issue when it was determined that Lovaas behavioural therapists were not classified as health care practitioners.

The next legislative framework - the BC Medicare Protection Act (MPA) - guides the administration of health care services. The purpose of the Act is, “to preserve a publicly managed and fiscally sustainable health care system for British Columbia in which access to necessary medical care is based on need and not an individual’s ability to pay” (Government of BC, 2004, p.1). It enshrines the principles of the Canada Health Act while attesting to the conviction that, “the people and government of BC believe that medicare is one of the defining features of Canadian nationhood” (BC Health Coalition, 2005). It acknowledges, “the judicious use of medical services in order to maintain a fiscally sustainable health care system for future generations” (Government of BC, 2004, p.1) and prohibits patients from being charged for a benefit or for “materials, consultations, procedures, use of an office, clinic or other place or for any other matters that relate to the rendering of a benefit” (Health Canada, 2006, p. 11).

The Medicare Protection Act establishes and regulates the British Columbia Medical Services Plan. Qualified residents are entitled to receive health care benefits with the payment for such services made to the service provider by the government. The MPA differentiates between services provided by a “medical practitioner” and those provided by a “health care practitioner” alongside the distinction made in the Canada Health Act between core and non-core services. “Medical practitioners” deliver core services and include physician services along with services provided within hospitals; these benefits must be fully funded and available to all residents. Core services can also be supplemented by partially funded, non-core services at the discretion of the Province. These include services provided by “health care practitioners” such as chiropractors, dentists, optometrists, podiatrists, and by regulation, physical therapists, massage therapists, and nurses. Therefore, a service cannot be a benefit under the Medical Services Plan unless it is provided by a medical practitioner or by a health care practitioner listed in section 13 of the Medicare Protection Act or in a regulation (Health Canada, 2006; Supreme Court of Canada, 2004a).

The Medical Services Commission is the regulatory agency charged under the Medicare Protection Act with implementing the Medical Services Plan. The Commission has the authority
to determine whether a service is considered a core benefit and whether a specific individual is deemed either a medical practitioner or a health care practitioner. Section 15(2) of the Medicare Protection Act also grants the commission power to cancel a practitioner's enrolment. The Commission does not have the authority to provide funding for a service not provided by a medical practitioner (e.g., non-core services). In these situations, a special request for funding is made to a Supplementary Practitioner Special Committee, specific to the type of health care practitioner administering the service. However, neither the Commission nor the Supplementary Practitioner Special Committee has the authority to fund non-core benefits if the proposed provider is not already listed in s. 13 of the Medicare Protection Act.

Judicious funding of non-core benefits is in compliance with the principles of the Canada Health Act. Provinces are free to limit or restrict benefits in terms of cost and number of treatments (e.g., restrictions on the number of insured eye exams); no service by a "health care practitioner" is fully insured. In addition, although the Medicare Protection Act requires that a service be "medically required", as previously mentioned, this term is not defined.

The inspiration for this thesis originated from the Auton legal proceedings that played out in both the trial and appellate Courts of British Columbia and the Supreme Court of Canada. The following sections will provide the background of these trials.
Legal Context

**Auton v. British Columbia (Ministers of Health, Children & Families, Education)**

*(Supreme Court of BC, Mar 31, 1999)*

In May 1998, on behalf of her infant son, Michelle Auton launched an application to serve as the representative plaintiff for autistic children in BC in a class action suit launched against the Province for failure to fund Lovaas Autism Treatment (LAT). The claimant cited a violation of s. 15(1) of the Canadian Charter of Rights and Freedoms (to be discussed in a future section). Section 15 of the Charter is frequently named in legal applications as it speaks to the right to equality: "every individual is equal before and under the law" (Morris, 1996, p. 279). More specifically, s. 15 prohibits discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability" (pg. 279). In the *Auton* case the claim was based on mental disability. The plaintiff was seeking relief for past and future costs of care incurred for LAT. The common issues in the class action suit identified by Madam Justice Marion Allan (Allan J.) were:

1. Whether Lovaas was an appropriate form of autism treatment.
2. And, if so, whether the plaintiffs were entitled to funding (both past and future).
3. Whether the defendants failed to exercise discretion by refusing funding for Lovaas.
4. Whether the defendants' failure to fund Lovaas violated s. 15(1) of the Charter.
5. Whether the defendants' failure to include Lovaas as an insurable benefit under s. 1 of the Medicare Protection Act violated s. 15(1) of the Charter.

The defendant Ministries of Health, Children & Families, and Education opposed the class action application and instead proposed that the plaintiff's case be heard via a judicial review process. A judicial review is, "a review by a Court of law of some act, or failure to act, by a government official or entity, or by some other legally appointed person or organized body" (Gifis, 2003, p. 278). Allan J. agreed with the government and dismissed the class proceeding application.

**Auton et al. v. Attorney General of BC and the Medical Services Commission of BC**

*(Supreme Court of BC, July 26, 2000)*

The objective of this subsequent proceeding was to determine liability in regard to the Charter violation claim. If the Crown was found to be at fault, appropriate remedies were to be addressed in a future hearing. Three additional children were added to the petitioners' original claim in order to represent the full spectrum of autism. The claim sought the following relief: (a) to receive a declaration from the Crown that failure to fund Lovaas contravened s. 7 and s. 15 of the Charter, and (b) entitlement to funding for all past and future Lovaas expenses. While, s. 15 of the Charter deals with discrimination, S. 7 states: "everyone has the right to life, liberty and security of the person and the right not to be deprived thereof except in accordance with the principles of fundamental justice" (para. 84) (also to be discussed in a
future section). The Crown responded by claiming that the children were not discriminated against. However, the Crown also stated that if there were a violation, it would be justifiable under s. 1 of the Charter (Section 1 provides Courts with the authority to potentially limit contested rights and freedoms).

One of the central issues of this case was the Lovaas efficacy debate. The petitioners argued that with Lovaas treatment many autistic children show improvement in language, socialization and intellect. On the other hand, the Crown contended that the efficacy of Lovaas had yet to be established and as such should not be ruled a “medically necessary service” provided under the Medicare Protection Act. Both counsel drew upon autism experts and published research studies on Lovaas therapy. In addition, reports were commissioned to support each party’s position. On behalf of the Crown, the British Columbia Office of Health Technology Assessment (BCOHTA) conducted one such study (Bassett et al, 2000). This health technology assessment study performed a systematic review and critical appraisal of the evidence on Lovaas therapy and concluded that, due to identified methodological flaws, there was a lack of scientific evidence to support the effectiveness claim of this treatment option. The petitioners, on the other hand, countered by stating that the gold standard of randomized controlled trials in autism was often not feasible due to a lack of resources. They cited expert opinion that the “quasi-random assignment” methodology utilized in these studies essentially accomplished the same outcome as random assignment (Baer, 1993). In addition, the petitioners funded a cost-benefit analysis of implementing a Lovaas treatment program (Hildebrand, 1999). In the end, the judge discredited the BCOHTA report, stating: “[that it] exhibit[ed] an obvious bias towards supporting the Crown’s position in this litigation that detract[ed] significantly from its usefulness” (para. 48). In addition, the Court concluded that it was beyond its jurisdiction to direct government to specifically fund Lovaas; therefore, the efficacy debate was viewed as being purely academic and irrelevant to the outcome of the proceedings. The judge also dismissed the petitioner’s cost-effectiveness report stating that it was impossible to estimate both the immediate costs of implementing a Lovaas program and the “inevitable savings” (para. 145) to the government in the future.

The Court ruled that the Crown had failed to meet its constitutional obligation of providing a “medically necessary service” (para. 102) and in doing so had contravened s. 15 of the Charter (paras. 139 & 158). This violation was also deemed unjustifiable under s. 1. The judge commented that, “the respondents’ argument that they are unable to provide effective treatment for autism because of constraints in the legislation governing medicare attempts to erect a false barrier” (para. 154). In fact, the judge proffered two potential funding models by which to comply with the ministries’ constitutional obligations. The Court’s first suggestion was to include EIBI in the MSP program and add behavioural therapists to the scheduled list of health
care providers; the second suggestion was to pay for the treatment through block funding (para. 154).

**Auton et al. v. Attorney General of BC**

(Supreme Court of BC, Feb 6, 2001)

At the subsequent remedy hearing, Allan J. directed the government to fund generic Early Intensive Behavioural Intervention (EIBI). In addition, since the Ministry of Children and Families, being a social services ministry, lacked the mandate and the expertise to deliver this treatment, the Court ruled that the program should be delivered through the Ministry of Health. The specifics of the petitioners' claim for remedy under the Charter included the following:

1. A declaration from the government that they breached the children's equality rights.
2. An order of mandamus (such an order would compel the government to abide by their statutory duty by initiating actions to rectify their Charter violation).
3. Financial compensation for all past and future expenses for Lovaas treatment (future treatment duration and intensity to be determined by either a medical practitioner or psychologist).

While the main points of the Crown's submission included the following arguments:

1. Their recommendation that the only remedy should be a declaration of their Charter violation, in addition to the provision of funding for EIBI therapy for children with autism and autism spectrum disorder. In other words, they opposed the following petitioners' requests:
   a. Funding EIBI "whenever, and to the extent that, a medical practitioner or psychologist recommends it" (para. 4). (Citing the reason that unlimited access to treatment would restrict policy options with respect to diagnosis and assessment).
   c. An order of mandamus (as the Crown felt the proposed P-CARD program was evidence of their intent to right a constitutional wrong.)
   d. Funding for EIBI past the age of 6 (stating that the age limit is based on evidence, the Reasons for Judgment, and the petitioners' own claim as to the effectiveness of treatment for young children [para. 39]).

In addition, the Crown also sought a delay, (a "remedial transition period"), to submit the declaration of a Charter violation in order to provide sufficient time to implement the P-CARD program. However, the petitioners challenged the validity of this program. They argued that the Crown had not revealed the specific treatment protocols and staffing qualifications, and that the program itself was nothing more than a scaled-up version of an earlier pilot project. They disapproved of the age restriction of 6 years (instead of upon the discretion of a physician); the amount of therapy set at 20 hours/week (instead of 40 hours); assessment and diagnosis...
performed by a multidisciplinary team (as opposed to a physician or psychologist); and the lack of a Lovaas component to the program (para. 24).

In her decision, Allan J. concluded that:

1. The Court still could not direct government to fund or provide a specific treatment, nor could the Court dictate how to allocate scare health care resources.
2. The evidence showed that Lovaas had been incorporated into other provincial, U.S. and international autism programs and was supported by many physicians in BC.
3. The Court still could not direct the government to provide EIBI, based on the advice of a physician or psychologist (as opposed to a multi-disciplinary team).
4. The government was obligated to provide effective treatment to autistic children to alleviate their disadvantaged state; this treatment could not be delayed or denied due to cost.
5. There was evidence to support the claim that early effective treatment of autism would result in large cost savings to the government in the long term.
6. The judge was unable to rule on whether the government breached its obligations if it failed to fund EIBI past the age of 6 years.
7. An order of mandamus was not warranted as the government was showing intent by implementing the P-CARD program.

Therefore, the Court ruled that (paras. 65-67):

1. The Crown must provide a declaration to the petitioners that its failure to provide effective treatment was a violation of their rights under s. 15 of the Charter.
2. The Crown must fund EIBI.
3. Each of the adult petitioners was entitled to “symbolic damages” of $20,000 (The judge determined that an order to compensate the petitioners for all past and future Lovaas expenses could set a precedent for all subsequent petitioners). Symbolic damages were not based on costs incurred (or to be incurred in the future); it merely recognized the role these petitioners undertook on behalf of all autistic children in BC.
4. Proceedings against the Medical Services Commission be dismissed, as it was the government that was in violation of the Charter, not the MSC.
5. The petitioners should be financially compensated for the costs of litigating the Charter claim.

**Auton et al. v. Attorney General of BC**

* (Court of Appeal for BC, Oct 9, 2002)

At the Court of Appeal, although the Crown did not appeal the symbolic financial award, they did appeal the s. 15 Charter violation ruling. They also contended that, should the appellant Court agree with the lower Court’s finding, any breach would be justified under s.1 of the Charter.
Reciprocally, the petitioners cross-appealed the remedy stating that the symbolic amount was insufficient to cover their damages. They also cross-appealed the generic nature of the EIBI order because they were specifically seeking Lovaas therapy.

Upon review of the case, one dissenting judge (Justice Lambert) disagreed with the lower Court's ruling on several points. With regard to the financial remedy, Justice Lambert rejected the symbolic award amount, instead proffering a formula that would best address the actual costs incurred by the plaintiffs. Pertaining to the intensity and duration of the treatment, Justice Lambert disagreed with the age cut-off for Lovaas at six years. Instead, he recommended an age limit of 14 years. In addition, Lambert J.A. felt that instead of the Crown's proposal for a multi-disciplinary team assessment, treatment should be funded upon written documentation from the family physician along with either a neurologist or a psychologist; these medical practitioners would be the ones to determine the specific type, intensity and duration of the program.

However, in the end, two of the three appellant Court judges (Justices Hall and Saunders) upheld the lower Court's ruling in regard to the Crown's Charter violation and the choice of funded treatment option (generic EIBI). However, they concluded that, since Lovaas had proven effective for the four child petitioners, the Ministry of Health should provide future funding for LAT for these specific children only (irrespective of their age, provided that it was still determined to be effective by a medical practitioner). They also dismissed the petitioners' cross-appeal. However, instead of a thorough assessment of the factors that may have contributed to a s. 1 claim, the majority of the Court invoked its parens patriae jurisdiction (a duty bestowed on the Courts to protect the rights of those perceived to be disadvantaged, e.g., children). The Crown subsequently applied for and was granted leave to appeal to the Supreme Court of Canada.

**Anderson et al. v. Attorney General of BC**

(Supreme Court of BC, 2003)

In the Auton trial, Allan J. reasoned that there was a good chance that a declaration of unconstitutionality in favour of the plaintiffs would produce significant changes in government policy that would indirectly benefit other autistic children whose rights had been violated. However, since the eventual outcome of Auton saw funding for LAT for only the four child petitioners (and generic EIBI for the rest of the autistic children), the petitioners from the denied Auton class action suit launched their own Charter challenge. In this trial the petitioners sought the same relief awarded to the plaintiffs in Auton. However, the Auton Court of Appeal judgment was not a "consolidated order" (para. 24) (e.g., inapplicable to all autistic children in BC).

In Anderson, Justice Pitfield ruled that the Crown had also violated these petitioners' rights and, since the Court deemed it unnecessary to litigate the same issues any further, he
declared that the petitioners were also entitled to LAT. However, Pitfield J. dismissed their claim to the $20,000 symbolic award.

The Attorney General of BC subsequently appealed the Anderson decision while the child petitioners cross appealed (Court of Appeal for BC, 2004). The Crown requested an extension of the deadline to file their appeal factum until after the outcome of Auton’s appeal to the Supreme Court of Canada. They argued that the two cases were intertwined and Auton’s outcome would impact their case. Levine J.A. denied the Crown’s request based on the child petitioners’ contention that, since the Attorney General had historically elected to litigate each autism case individually, there would be no reason to delay the appeal. However, the Auton Supreme Court of Canada ruling was handed down before the Anderson appeal was heard. Therefore, in February 2005, on the initiative of the Crown, the Court of Appeal for BC allowed the Crown’s appeal (and dismissed the petitioners’ cross appeal), thereby setting aside the lower Court’s ruling that granted Lovaas treatment to the remaining children from the original Auton case.

(Supreme Court of Canada, 2004a)

On June 9-10th, 2004, the Supreme Court of Canada heard the Auton appeal. In addition to submissions from the Attorney General of BC and counsel for Auton, the Court heard arguments from various interveners from across the nation such as other provincial Attorneys General, along with various social actors and autism groups such as Women’s Legal Education and Action Fund (LEAF), the DisAbled [sic] Women’s Network of Canada (DAWN), and Families for Early Autism Treatment (FEAT). Each stakeholder spoke to specific findings of fact from the lower Courts, and/or conceptual and constitutional inconsistencies.

The Attorney General of BC’s submission focused on whether the respective definitions of “benefits” and “health care practitioners” found in the BC Medicare Protection Act infringed on the petitioner’s s. 7 and s. 15 rights as set out in the Charter. The Crown’s position was that their funding decision was not discriminatory and that “a decision to delay, refuse, ration, or experiment with public funding for a new treatment [was] not a decision about the worth or dignity of those suffering the disabling condition” (Supreme Court of Canada, 2004b [argument of Cowper, G. – counsel for the Attorney General of BC]). It was further contended that a judicial process essentially usurped an essential governmental policymaking framework that takes into account a myriad of polycentric values. One of the Crown’s central arguments was the fact that the Supreme Court of BC determined the therapy was efficacious and in doing so, inferred that a delay or refusal to provide an efficacious treatment constituted discrimination. The Crown did not contest that “behaviour therapy” was effective; the issue revolved around the inference of discrimination. In fact, the Crown admitted that their funding decision may have been “wrong, cheap, unwise, or late” (Supreme Court of Canada, 2004b [argument of
Cowper, G.) but not discriminatory. They argued that the nature of the Auton case was for funding, not equality.

Another main argument presented by the Crown pertained to the lower Courts' error in judicial reasoning in regard to the Charter discrimination analysis. This error was seen as being perpetuated by Allan J.'s incorrect application of the Canada Health Act's principles of universality and comprehensiveness:

The exclusion of effective treatment for autistic children undermines the primary objective of the Medicare legislation, which is to provide universal [italics added] health care. (Supreme Court of BC, 2000, para. 151)

As previously mentioned, in the Canada Health Act, universality refers to access to the health care system, whereas comprehensiveness refers to specific services delivered by physicians and dentists within hospitals. The Crown pointed out that there was no constitutional or statutory right to efficacious treatment. In fact, they warned the Supreme Court that a ruling in support of a violation of s.15 of the Charter would in effect create a "constitutionally-mandated right to public funding" for all efficacious programs.

The submission by Chris Hinkson, counsel for Auton, highlighted the disputed facts of the case. First of all, he noted that the Crown did not appeal Allan J.'s ruling that the Lovaas issue resided within the scope of the Ministry of Health's jurisdiction (as opposed to the Ministries of Education or Children and Families). Second, the Crown did not initially approach the case using a financial argument (as he contested they were now doing). Instead, the Crown's strategy had been to debate the Lovaas efficacy claim. However, Mr. Hinkson argued that it would cost between thirteen and thirty-three million dollars per year not to treat autism in BC. Third, Chris Hinkson stated that there was no evidence that the Crown weighed the effectiveness of Lovaas treatment, or entered into a cost-effectiveness analysis. This is an interesting statement in view of the fact that the Crown had commissioned the BCOHTA report, which was given significant attention by Allan J. in her reasonings. And finally, Mr. Hinkson contended that there was no evidence that supported a systematic and transparent approach to this policy option on the part of the government. To this remark Justice Binnie proffered a suggestion that an appropriate order from the lower Courts might have been to direct the government to conduct a proper policy analysis in order to determine entitlement.

Some of the submissions from the interveners in the case presented slightly varied perspectives. For example, the Autism Society of Canada (ASC) had a more systems approach, focusing on government's lack of a "constitutionally appropriate process and criteria to support its decision not to fund Lovaas" (Supreme Court of Canada, 2004b [argument of Crolla, D.A. & O'Brien, M.K.]) They stressed the need for accountability and transparency in policy development. The Canadian Association for Community Living (CACL) and the Council of Canadians with Disabilities (CCD) on the other hand reasoned that a comparator group analysis
was not appropriate for a s.15 Charter analysis. Instead, they felt that the pressing question was: “Did they [autistic children] get what they needed?” (Argument of Shilton, E.J.; Faraday F.C. & Chadha, E.) Their position was that Courts needed to supervise the frameworks that governments use for resource allocation based on equality rights; they argued that current decision-making processes do not include an impact analysis of these rights.

FEAT of Ontario and Alberta argued that a critique of the Auton case written by Donna Greshner and Steven Lewis (2003) was in fact a “discriminatory report”. This paper, which was frequently referenced by various interveners in this case, examined the claims made in Auton and the judicial reasoning of the lower Courts that resulted in policy being dictated by law. FEAT felt that the Auton case was not about establishing a constitutionally guaranteed right to health care but rather ensuring access to a core, medically necessary and publicly funded health treatment for vulnerable autistic children.

An interesting submission was presented on behalf of an adult with autism - Michelle Dawson. Her counsel stated that she did not support either side in the case and resented the assumption that autistic children required treatment in order to be “normal” (Ms. Dawson had refused EIBI/ABA therapy). In addition, her lawyer disputed the often stated statistic of only 1:64 children achieving “normal functioning” without some form of intervention stating that the research was published in 1970. Their argument centered on the fact that neither party had taken the time to consult an autistic person.

On the respondents’ side, counsel for the children claimed that the government was funding medically necessary treatment for non-disabled Canadian children and adults with mental illness, yet autistic children were being denied the same benefit. McLachlin CJC found that such an assertion equated to funding for all medically required treatment.

On November 19, 2004 the Supreme Court of Canada handed down its judgment (Supreme Court of Canada, 2004a). The Court allowed the appeal, and ruled that the BC government’s conduct had not infringed upon the respondents’ equality rights under s. 15 of the Charter. It found that both the Canada Health Act and the Medicare Protection Act failed to support the respondents’ assumption that all medically required or necessary non-core services must be funded. Instead, the Court pointed out that the Acts confer only funding for core services provided by medical practitioners; any funding for non-core services was at the discretion of each Province. The Acts were not intended to meet all medical needs of the people; in fact, McLachlin CJC called the legislative scheme “a partial health plan” (para. 43). However, some non-core services can be funded through the Medical Services Commission (MSC) of BC (the administrative body that governs the Medicare Protection Act). Nevertheless, at the time of the original trial, behavioural therapists were not a recognized health care practitioner, nor did the MSC have the authority to create this classification. Therefore, ABA/IBI therapy could not be funded under the plan. Consequently, the Court’s ruling was that BC’s law
governing non-core benefits did not support the respondents' case. Therefore, based on this analysis, the Supreme Court held that funding for all medically required services was not a constitutional right provided for by the law.

A point of interest in McLachlin's s. 15 analysis was her speculation on the outcome of the inquiry if the respondents had framed their case differently: "as a claim to the benefit of equal application of the law by the Medical Services Commission" (para. 45). Her hypothesis being that the health care scheme itself was discriminatory as it provided funding for some non-core therapies while refusing to fund an "equally necessary" therapy such as ABA/IBI. However this claim would have been contingent on showing that there was a benefit provided by law. This means that, at the time of the case, there needed to be a non-core service (the "benefit") being provided under legislation. Therefore, if behavioural therapists had been listed as health care practitioners then the Medical Services Commission would have been obliged to provide LAT.

Another issue central to the s. 15 analysis was the choice of a comparator group. Here the respondents had to prove that they were denied a benefit that was typically available to a similar group of people. The respondents chose their comparator group to be non-disabled children and their parents, along with mentally ill adults. However, the Court disagreed and decided that a more appropriate comparator group would be a "non-disabled person or a person suffering a disability other than a mental disability . . . seeking or receiving funding for a non-core therapy important for his or her present and future health, which is emergent and only recently becoming recognized as medically required" (para. 55). The Court acknowledged the "emergent" nature of Lovaas therapy and decided, "people receiving well-established non-core therapies are not in the same position as people claiming relatively new non-core benefits" (para. 55) and that governments may have legitimate reasons for delaying or denying funding. McLachlin CJC pointed to the fact that evidence on how the government responded to requests for non-core services from non-disabled or otherwise disabled people would have been beneficial in determining whether the government discriminated against the child petitioners, or whether it proceeded similarly with all requests. Based on this lack of evidence, the Supreme Court of Canada could not support the claim of discrimination.

And finally, the petitioner's also claimed a violation of s. 7 of the Charter. However, McLachlin CJC pointed out that the respondents did not identify the principle of fundamental justice that had been contravened by the refusal to fund Lovaas therapy. Principles of fundamental justice are generally agreed upon legal principles that are "vital or fundamental to our societal notion of justice" (Supreme Court of BC, 1993, para. 590). Therefore, based on a lack of evidence, there was no validity to claim a s. 7 violation.
Again, to better understand the context in which the Auton and Anderson trials played out, an in-depth look at the Canadian legal system and some of its relevant statues, along with some relevant cases, is required.

Table 1.1 below provides a summary of the pertinent legal cases.

<table>
<thead>
<tr>
<th>Case</th>
<th>Issues</th>
<th>Judgments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auton (Mar 31, 1999) Supreme Court of BC</td>
<td>Class action application against BC government for failure to find Lovaas Autism Treatment (LAT) (Charter challenge)</td>
<td>Class action denied. Case proceeded via judicial review.</td>
</tr>
<tr>
<td>Auton et al. (July 26, 2000) Supreme Court of BC</td>
<td>Determination of liability in regard to Charter violation.</td>
<td>Province found in violation of the Charter for not providing a &quot;medically necessary&quot; service. Ordered to fund Early Intensive Behavioural Intervention (EIBI).</td>
</tr>
<tr>
<td>Auton et al. (Feb 6, 2001) Supreme Court of BC</td>
<td>Claim for remedy hearing. Parents sought funding for all past and future LAT expenses.</td>
<td>Parents only granted $20,000 as a &quot;symbolic award&quot;.</td>
</tr>
<tr>
<td>Auton et al. (Oct 9, 2002) Court of Appeal for BC</td>
<td>Province appealed the Charter s.15 violation. Petitioners appealed the $20,000 remedy and the generic nature of EIBI services.</td>
<td>Lower court rulings upheld. However, LAT awarded to the four child petitioners.</td>
</tr>
<tr>
<td>Anderson et al. (Aug 22, 2003) Supreme Court of BC</td>
<td>Similar case brought against the province by the remaining children left out of the original Auton proceedings.</td>
<td>Province again found in violation of the Charter. Children granted funding for LAT, however did not receive the $20,000 &quot;symbolic award&quot;.</td>
</tr>
<tr>
<td>Auton et al. (Nov 19, 2004) Supreme Court of Canada</td>
<td>Whether a provincial health plan is obligated under the Canada Health Act to provide treatments outside of the core services, and if so, whether failure to fund these services constituted a violation of the Charter.</td>
<td>Province of BC won their appeal. Court ruled that the Canada Health Act and the Medicare Protection Act does not provide a legal right to all medically required or necessary non-core services.</td>
</tr>
</tbody>
</table>
Canadian Legal System

Canada's judicial system (with the exception of Quebec) is derived from the English structure, which is based on judge-made law called common law. During the early 11th century the English government appointed judges to settle disputes. Lacking formal laws upon which to guide their decisions, these early cases were tried based on custom, tradition, societal norms and common sense. These historical judgments set precedent for future trials thus providing the basic legal principles, concepts, language, methods of legal reasoning, and divisions of law for the modern day legal system. Since confederation in 1867, Canadian laws have evolved principally from federal parliamentary and provincial legislative laws called statutes. The Constitution Act of 1982 (Canada Act) granted authority to the Provinces to enact their own statutes within the legislative scope defined in the Act. Statutory laws prevail over common law (Sneiderman, et al, 1995). In the common law system, either the judge or jury is charged with the determination of facts and the application of the relevant law to those facts. If a trial is by jury, the judge is responsible for determining the legal issues while the jury decides the factual matters. If the trial is by judge alone, their role is to provide objective arbitration over specific disputes. Objective arbitration is a process that initially seeks out the truth and then applies specific rules of procedure and rules of evidence in order to determine the outcome of the case (Gall, 2004).

Health law is a relatively new discipline that has risen from tort law. The law of torts is a branch of common law that deals with injuries or damages inflicted upon a person by another. A number of factors have contributed to health law becoming a discipline unique unto itself. These include the scientific explosion in areas such as genetic testing and biotechnology; complex ethical dilemmas brought about by end of life issues such as euthanasia; unsatisfactory patient care outcomes as a result of health care reform policies; and privacy issues over the advancement of information technology, just to name a few (Downie & Caulfield, 1999).

The function of the Canadian Courts is, "to provide fair and just resolution of the various problems and conflicts that are brought before them" (Gall, 2004, p. 209). This judicial decision-making takes place within the context of the adversarial system where the assumption is that the truth is most likely to be uncovered. One of the key features of judicial decision-making is the process of adjudication. The role of adjudication is in settling disputes between private individuals or groups, or the government and the individual. Usually these are not future debates of general policy issues; instead, they are immediate controversies that occur when opposing interests intersect. Gall (2004) describes the fundamental objectives of the Canadian judicial system as the following:
Our Courts must entertain a search for truth, and that search for truth must be conducted in a manner and with the result that might, broadly speaking, be characterized as the dispensation of justice. In turn, the dispensation of justice must not only be directed at those persons appearing before our Courts, but also at the same time it must be directed to the best interests of society at large. (p. 212)

The Canadian judiciary is divided into a hierarchy of Courts each with their own unique jurisdictional responsibilities. Provincial Courts are governed by their respective provincial statutes and consist of Courts of inferior jurisdiction (e.g., provincial Courts) to Courts of superior jurisdiction (Supreme Court and Court of Appeal). However, while Provinces appoint lower provincial Court judges (e.g., Youth Court, Family Court, Small Claims Court and the criminal jurisdiction of the Provincial Court), the higher Courts are under a federal constitutional limitation whereby judges must be federally appointed. While each Province shares similar fundamental objectives, the manner in which they are constituted varies among Provinces. This results in different responsibilities for apparently similar Courts across the country.

Procedures are in place whereby parties can dispute the outcome of trials. If in counsel’s opinion the judge has erred in either the finding of fact or in the application of law the option to appeal the decision is available through the provincial Court of Appeal. A panel of the Court of Appeal usually consists of three judges; majority decides the outcome of the appeal. Although error in fact is grounds for appeal, the mandate of the appellant Court is primarily to judge the application of law to the facts as outlined in the Reasons for Judgment. The unsuccessful litigant seeking to over-turn the trial decision is called the appellant while the successful party is termed the respondent. If the successful party also disputes the judgment (for example, asserts that the financial settlement was not sufficient) they are called the appellant on cross appeal and the other party is the respondent on cross appeal (Sneiderman, et al., 1995).

The doctrine of precedent dictates that once a higher Court rules on a specific issue the lower Courts must apply the same ruling (provided that the facts of the case are similar in nature). Hence, depending on the position each counsel is taking, the lawyers’ mandate is to carefully examine the facts of precedent cases to determine whether they apply to their case before the Court or whether there are distinguishing facts that would render the doctrine of precedent not applicable (Sneiderman, et al., 1995).

The federal Courts consist of the Federal Court of Canada, the Tax Court of Canada and the Supreme Court of Canada. These Courts are governed by federal statues with federally appointed judges. The Supreme Court of Canada is the highest Court in the country and functions as the ultimate appellate Court for the Court system. The Supreme Court is comprised of nine judges (one Chief Justice of Canada and eight “Puisne Justices”) that preside in an odd number in order to prevent decisions being split evenly. The number of judges that hear the proceedings is based on the severity of the case (Sneiderman et al., 1995). Between 1984 and
1997 the Supreme Court of Canada nullified thirty-four federal and seventeen provincial statutes (Manfredi & Maioni, 2002).

**Canadian Charter of Rights and Freedoms**

The Canadian Charter of Rights and Freedoms was enacted in 1982 as part of the "new" Constitution of Canada of the Canada Act. The Constitution of Canada is the supreme law of the nation. The Charter replaced the 1960 Bill of Rights, which at the time was powerless over provincial statutes as it was not a constitutional document. The Charter contains essentially the same elements as the original Bill of Rights, but individuals' rights and freedoms were more clearly codified. In addition, the Charter is applicable to both federal and provincial levels of government and is intended to protect the people from arbitrary or illegal legislation. If a statute appears to contravene the Charter it can be challenged in the Courts and potentially deemed unconstitutional (Sneiderman, et al, 1995). The only exception is found in s. 1 "where the government can show reasonable and demonstrable justification in a free and democratic society" (Mykitiuk & Wallrap, 1999, p. 320). Therefore, rights and freedoms are not absolute, and Courts are charged with weighing the claim against all other competing societal interests. Accusations of violation of the Charter are but one mechanism by which plaintiffs may seek access to specific health care services.

Between 1984 and 1997 the federal government was able to defend against Charter challenges in 62.5% of the cases, while provincial governments have been successful 72.3% of the time. The nullified provincial statutes were significantly 'younger' than the annulled federal statutes with a mean 'age' of ten years as opposed to twenty-three (Manfredi & Maioni, 2002).

**Section 15**

Section 15 of the Charter is frequently cited in legal applications since it speaks to equality rights: "every individual is equal before and under the law" (Morris, 1996, p. 279). More specifically, s. 15 prohibits discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability" (pg. 279). S. 15 claims in health care are typically focused on expanding the scope of insured services. In these cases, patients will allege that a specific health care policy either excludes them from coverage or reduces their share of resources, thus violating their rights to equal protection, and equal benefit of the law. While the potential for s. 15 claims is large, in actual fact only a few cases have been successful.

Since Auton represents a scope of coverage case, in order to prove a violation of equality rights, (under the steps established in Law v. Canada [1999]), the plaintiff must prove three things: (a) differential treatment between groups based on a personal characteristic; (b) the differential treatment must be on a ground that is enumerated (powers granted by the Constitution) in s. 15 or analogous to an enumerated ground; and, (c) the differential treatment must be so substantive that it offends the plaintiff's essential human dignity (Greschner, 2002).
However, rationing of health services due to cost, risk, safety and low effectiveness would not constitute a violation of rights.

**Section 7**

Section 7 of the Charter states that, "everyone has the right to life, liberty and security of the person and the right not to be deprived thereof except in accordance with the principles of fundamental justice" (Supreme Court of BC, 2000). However, proving this in the health care context has been less successful. A s. 7 claim requires that the plaintiffs initially show that either "liberty" or "security" incorporates health or health care services. For example, Rodriguez v. British Columbia (a case on euthanasia) was successful in arguing that an individual's right to refuse medical treatment fell within the Charter's jurisdiction of "security". Then, once the plaintiff has established the liberty/security criteria, they must prove that the deprivation of the right to liberty and security contravened the principles of fundamental justice; this means that the basic tenets of the legal system would have to have been broached. S. 7 challenges in health care are incredibly hard to prove as this rule typically applies more to criminal cases.

**Section 1**

And finally, s. 1 of the Charter contains the "limitations clause", which provides Courts with the authority to potentially limit contested rights and freedoms. The framework for judicial reasoning and analysis of a s. 1 claim is referred to as the Oakes test as the criteria for supporting the limitations clause was decided in R. v. Oakes (1986). S. 1 allows governments to justify violations to the Charter as reasonable limits. In order to do this though, two criteria must be satisfied: (a) "the impugned law must pursue an objective that is sufficiently important to justify limiting a Charter right", and (b) "the government must demonstrate that the means chosen to attain it are reasonable and demonstrably justified" (Hartt & Mohahan, 2002, p. 23). Satisfying the first criteria is not difficult as governments can argue that their policy objective is to protect the Medicare system by effective use of resources. However, the second criterion is somewhat harder to demonstrate and requires clear evidence that other policy options have been considered, that the policy impaired the right or freedom as little as possible, and that it should not disproportionately effect the targeted population (Greschner, 2002). In essence, the s. 1 requirement of demonstrability resembles the evidence-based approach taken by policy-makers. In fact:

In the majority of Charter cases Courts perform their most important task not in defining the substantive meaning of rights or liberties, or in measuring government action against those definitions, but in determining the scope of "reasonable limits" on rights under s. 1 of the Charter. (Maioni & Mantredi, 2003, p. 928)
Other Health Care Charter Cases

While Auton certainly captured the media’s attention, some other health care cases that both preceded and followed Auton have also contributed to this ongoing debate about judicial policymaking and Charter challenges being used to access health care services.

*C.R. v. Alberta (Director of Child Welfare) (Alberta Court of Queen's Bench, 1996)*

**Facts and Issues**

This case was the first to consider funding for treatment for autistic children. However, as opposed to determining what type of treatment to fund, it focused on which government department was accountable for delivering it. Here, the parents of an autistic child sought government funding for Lovaas therapy. In this particular situation, the child was diagnosed later than usual at the age of 5. He went on to receive Lovaas treatment beginning at the age of 6 for 20 hours per week outside of school hours. After personally funding the therapy for nine months, the family requested financial assistance from their school district. The application was refused based on the claim that the district was not responsible for supporting programs outside of normal school hours. Then, pursuant to the Child Welfare Act, the family approached Alberta’s Handicapped Children’s Services (a division of Alberta Family and Social Services). Once again, their request was denied citing the ruling that special education programs were within the realm of the School Act and as such superseded the Child Welfare Act. The parents proceeded to file legal action against the Director of Child Welfare claiming a violation of s. 15 of the Charter of Rights and Freedoms.

Three of the key issues for these proceedings were as follows: (a) whether the Director of Child Welfare erred when determining that Lovaas was under the exclusive jurisdiction of the School Act, (b) whether the Director of Child Welfare violated s. 15, and (c) if the Charter was violated, whether the Court should order funding pursuant to s. 83 of the Child Welfare Act.

**Reasons**

Justice Deyell ruled that the Director of Child Welfare had erred by characterizing Lovaas as an educational program; therefore, there was no justifiable reason to withhold treatment. In addition, depending upon the particular circumstances of each individual case, the judge concluded that Lovaas could also be appropriately funded under the School Act. The Court’s decision on this key issue made the question of a Charter violation irrelevant; therefore that claim was dismissed. Justice Deyell ordered the Director of Child Welfare to fund 90% of the costs of therapy for a one-year period and to reimburse the parents 90% for expenses incurred over the previous year. The judge decided that Lovaas therapy was to be funded by the Ministry of Children’s Services and not through health insurance. However, subsequent cases have declined to follow the legal reasonings in C.R. In one particular case (Alberta Court of Queen’s Bench, 1999) Justice Rawlins commented on these proceedings stating that the Courts should not have interfered with the government’s right to decide whether or not to fund specific
services, nor should it have ordered a certain percentage of the services to be funded (paras. 9 & 22). In any regard, the outcome of C.R. was that Alberta instituted a pilot project utilizing ABA interventions; Ontario, Prince Edward Island, Newfoundland and Manitoba all followed suit shortly thereafter with their own adaptations of ABA programs (Supreme Court of BC, 2000, paras. 73-81).

D.J.N. v. Alberta (Child Welfare Appeal Panel) (Alberta Court of Queen’s Bench, 1999)

Facts and Issues

In this case, D.J.N. appealed a decision by the Child Welfare Appeal Panel that struck down a request to provide financial assistance for a variety of additional programs for her autistic son B.N. These programs included speech therapy, occupational therapy, a Reading Foundation program, social skills and computer language training. B.N. was enrolled in a special education program, but was on the waiting list for some additional services such as speech therapy. The appellant claimed the child was unable to keep up with his peers, requiring services beyond what the school was providing.

The Director of Child Welfare denied these requests on the basis that it was not within his jurisdiction to grant these services. The appellant appealed the decision to the Child Welfare Appeal Panel who in turn upheld the decision stating that the requested services were under the jurisdiction of the School Board under the School Act.

The questions for the Court were: (a) Did the Panel fail to apply the law equally to B.N.? and, (b) Did the Panel err in finding it did not have jurisdiction over the requested services? (para. 7).

Reasons

Justice Rawlins found that the Panel did not violate s. 15(1) of the Charter, as the evidence on whether the appellant received inconsistent treatment was inconclusive. In addition, entitlement to services under the Child Welfare Act was at the discretion of the Director. And, in regard to the second question, the Court determined that the Panel had not erred in finding it did not have jurisdiction over the requested services.

Eldridge v. British Columbia (Attorney General) (Supreme Court of Canada, 1997)

Facts and Issues

This was the precedent case for all future health care Charter litigation. Here, three deaf appellants sought a declaration that the failure to provide sign language interpreters as an insured benefit violated s. 15 (1) of the Canadian Charter of Rights and Freedoms. At the time of litigation, neither the Hospital Insurance Act nor the Medical and Health Care Services Act (as it was known at the time) covered language interpreters. The appellants asserted that, “the absence of interpreters impair[ed] their ability to communicate with their doctors and other health care providers, and thus increase[d] the risk of misdiagnosis and ineffective treatment” (para. 3).
Eldridge rose to the Supreme Court of Canada in 1997 where it was debated whether the definition of “benefits” in s.1 of the Medicare Protection Act, in addition to various sections of the Hospital Insurance Act, infringed on s. 15(1) of the Charter. If the Court determined there was a violation, then the question remained whether it was saved by s. 1 of the Charter. In addition, there was the issue in regard to the extent that decisions made by private entities become subject to Charter review and the extent to which governments must provide equal access to public services for the disabled (Cornish & Faraday, 1997).

Reasons

The Supreme Court of BC ruled against the s. 15(1) Charter violation with Justice Tysoe’s reason being that “sign language interpretation [was] ancillary to medically required services in much the same way as [was] transportation to a doctor’s office” (para. 11). In his view, the Charter did not stipulate which programs a government should implement, only that the implemented programs should be distributed equally.

In 1995 the Court of Appeal for BC upheld the lower Court ruling stating that, “the lack of interpreting services in hospitals [was] not discriminatory because the Hospital Insurance Act [did] not provide any “benefit of the law” within the meaning of s. 15(1) of the Charter” (para. 13). The absence of interpreters was not caused by the legislation but by the hospital’s discretion as to how their global budget was distributed. And, since hospitals were not “government” as defined in s. 32 of the Charter their failure to provide these services did not constitute a violation of s. 15(1). The Court also found that the Medical and Health Care Services Act also did not violate s. 15(1) because there was no distinction between the deaf and hearing populations in that both populations had the benefit of free medical services.

On appeal to the Supreme Court of Canada (1997), the majority overturned the lower Courts' rulings and found that the government’s actions violated s. 15(1) of the Charter. They concurred that, “the evidence clearly demonstrate[d] that, as a class, deaf persons receive[d] medical services that [were] inferior to those received by the hearing population” (para. 94). By providing a benefit, the government had to ensure its accessibility to all. Even though hospitals are private entities, their responsibility to uphold the Charter is extended by way of their contract to implement government policies. Eldridge therefore broadened the range of entities and activities that could theoretically be subject to a Charter challenge.


Facts and Issues

George Zeliotis was a Quebec resident that sought out private health insurance due to delays he had experienced in receiving treatment through the publicly funded health care system. However, s. 15 of the Quebec Health Insurance Act and s. 11 of the Quebec Hospital Insurance Act prevented him from doing so.
Jacques Chaoulli was a physician who sometimes worked outside of the public health care system. He had applied to the Quebec Health Insurance Board to operate a private “opted-out” hospital but was denied. In addition, his medical practice of providing home visits on a 24-hour basis was also in question. His concern was that timely health care would not be available should he or his family require it. He, too, felt he should be able to purchase private insurance to cover medical services currently provided within the public system.

Chaoulli and Zeliotis brought a motion before the Superior Court of Quebec seeking a declaration that s. 15 of the Health Insurance Act and s. 11 of the Hospital Insurance Act violated ss. 7, 12 and 15 of the Charter in addition to ss. 1, 4, 5 and 24 of the Quebec Charter of Human Rights and Freedoms. However, the real issue that was identified by the Superior Court judge was the introduction of a private health care system parallel to the public system (Manfredi & Maioni, 2005; Supreme Court of Canada, 2005; Tiedemann, 2005a).

**Reasons**

The Superior Court of Quebec (2000) did not analyze the case in respect to the Quebec Charter. Instead, Madam Justice Piche identified the key issue as whether the provisions of the two provincial Acts violated the applicants' rights under s. 7 of the Canadian Charter. She determined that the right to obtain private insurance was an "ancillary or incidental economic right" (Tiedemann, 2005a, p. 4) since in this case it was linked to life, liberty or security of the person. She agreed that the Quebec Acts created a barrier to access required health services. However, a s. 7 violation of the Canadian Charter could only be declared if the public health system could not guarantee access to services, which was not the situation in this case. Therefore, no violation of s. 7 resulted and the Court ruled that the applicants' rights were not violated under either ss. 12 or 15 of the Canadian Charter.

The case proceeded to the Quebec Court of Appeal in 2002 where it was dismissed. All three Justices could not concur on whether the right was purely an economic right or an ancillary or incidental economic right.

And finally in 2005, four of the seven justices of the Supreme Court of Canada agreed that prohibiting the purchase of private insurance violated s. 1 of the Quebec Charter and that such a violation was not justified under s. 9(1). The Court, however, was equally split in regard to whether this prohibition violated s. 7 of the Canadian Charter since one Justice only analyzed the case from the perspective of the Quebec Charter. Madam Justice Deschamps concluded that the evidence did not support the hype surrounding the contention that allowing private insurance to cover services normally delivered by the public system would result in a parallel system that would bleed the publicly funded one. The dissenting justices were critical of the majority decision in that they felt the Court had not defined "how much health care is 'reasonable' enough to satisfy s. 7 of the Canadian Charter of Rights and Freedoms... and s. 1
of the Charter of Human Rights and Freedoms" (Supreme Court of Canada, 2005, para. 163) (Flood & Sullivan, 2005).

Barclay (Guardian ad litem of) v. British Columbia (Attorney General)
(Supreme Court of BC, 2005)

Facts and Issues

This action was initiated in March 2002 in response to the outcome of Auton. Patrick Barclay was an 11 year-old child with autism who did not receive any autism treatment until the age of 7½. At that time he began a program of applied behavioural analysis, paid for by his family, until such time as they were no longer able to afford treatment. This action also claimed a breach of s. 15 of the Charter and sought reimbursement for all past and future treatment.

This case never proceeded to trial as the Auton judgment from the Supreme Court of Canada determined the inevitable outcome of any relevant cases in progress. Instead, the plaintiffs sought an order from the Supreme Court of BC for special costs arising from the proceedings. In response, the defendants opposed the plaintiff’s application and in turn sought an order to cover their own costs. Some of the issues debated included: (a) whether the defendants were considered the successful party in light of the Supreme Court of Canada ruling and as such, their costs would be covered; (b) whether the plaintiff’s decision to proceed with the trial before the Supreme Court of Canada judgment in Auton would be considered a waste of time and money; and (c) whether the plaintiff was considered a “public interest litigant” (a party that has no personal outcome in a trial) (para. 16).

Reasons

The Supreme Court of BC found that the plaintiffs were entitled to their costs, while the defendants were denied theirs. The defendants subsequently appealed to the Court of Appeal for BC (2005).

R. v. Wynberg (Ontario Superior Court of Justice, 2005)

(Court of Appeal for Ontario, 2006)

Facts and Issues

This litigation represented 35 children from 30 families who were diagnosed with autism. It was based on access to early intervention therapy after the age of 6 as a right under the Ontario Education Act. The Ontario Superior Court heard the plaintiffs’ allegations that the Province’s Intensive Early Intervention Program (IEIP), which provided or funded treatment for children ages 2 to 5, violated s. 15 of the Charter as it discriminated against the children on the basis of age. They argued that their s. 7 rights were also violated. The plaintiffs were not challenging the constitutionality of any legislation or regulations; instead, they were challenging the government’s actions or inactions relevant to the design and implementation of the IEIP. In addition, one of the families (Deskin) also filed a negligence claim.
The government's response was that the age restrictions were based on financial constraints and were in no way intended to be discriminatory.

**Reasons**

Justice Kiteley ruled that the government had violated s. 15(1) of the Charter on the basis of age and disability and that this violation was not justified under s. 1. In addition, the Minister of Education was found to have violated his duty under s. 8(3) of the Education Act by failing or refusing to ensure the appropriate programs for school age children under s. 15 of the Charter. The Court denounced the same violation under s. 7 of the Charter and also denied the negligence claim by the Deskin family. Justice Kiteley also found that "IBI/ABA [was] not an 'emergent' therapy or treatment" (para. 27); this is particularly interesting in light of the earlier Supreme Court of Canada's assertion to the contrary (Auton, 2004). As a remedy, Justice Kiteley ordered a declaration of violation of the Charter in addition to financial compensation to the petitioners for past and future IBI/ABA treatment. The government of Ontario appealed the decision (Tiedemann, 2005b).

On July 7 2006, the Court of Appeal for Ontario struck down the lower Court's decision and unanimously ruled that the Ontario IEIP was not discriminatory and therefore did not contravene the Charter of Rights and Freedoms. In response to the claim of age discrimination, the Court found that although autistic children age six and older had been treated differently based on age, no evidence was presented to establish that this differential treatment constituted discrimination. Pertaining to the charge of disability discrimination, again, the Court of Appeal disagreed with the lower Court's finding of violation of the claimants' equality rights on the basis of disability. The Court of Appeal concluded that the claimants had failed to demonstrate that their clients had received differential treatment by not receiving the IEIP. The Court of Appeal observed that it had not been demonstrated that a program of this type could even be administered in a school setting (due to its intensity), nor did the claimants proffer evidence of the effectiveness of the existing programs for autistic children or for children with other disabilities. The Appeal Court's reasons being that this additional evidence would have been essential to establish that the IEIP was the only effective alternative and that by denying it would culminate in discrimination. The Court of Appeal also revoked the damages ordered by the lower Court on the basis that in general, damages are not afforded when declaratory relief is provided. In addition, they did not find legal ground for the s. 7 claim declaring that since the IEIP was not shown to be the only effective program for autistic children, there was no factual basis to conclude that it was fundamental to their personhood and development. The lawyer for the plaintiff was seeking leave to appeal to the Supreme Court of Canada.

Of particular interest in the ruling of the Ontario Court of Appeal was its s. 1 analysis (even though a finding of discrimination was not made). The Court of Appeal disagreed with the trial judge by stating that the age cut-off of 6 years was a reasonable limit under s. 1 of the Charter.
The Court acknowledged the government's role in allocating scarce resources to where the greater good could be done, and based on the evidence, this was decided to be in the autistic population under the age of 6. They reasoned that spreading limited financial and human resources over a larger age cohort would dilute the effectiveness of treatment to children under the age of 6, which it is purported to be the most efficacious "window of opportunity". The Court of Appeal directed lower Courts to consider program objectives and competing social interests when confronted with claims against allegedly under inclusive government programs. They also deferred to the expertise of the government when it comes to making difficult policy choices.

**Hewko v. Attorney General of BC and School District No. 34 (Abbotsford) et al.**

(Supreme Court of BC, 2006)

**Facts and Issues**

This case was centred on a school district's response to a child's disability and resultant special needs requirements to access a public education. Darren Hewko was a 9 year-old boy who was diagnosed with autism at the age of 3. He received home-based Lovaas Autism Treatment for two years until he started half-day kindergarten. At this point he continued to receive LAT at home while at kindergarten he was assigned a special education assistant. The school district permitted the Lovaas behavioural therapist into the school in order to transition Darren into the classroom. However, Darren's parents wanted the special education assistant to be a member of the home therapy program. In addition, they requested Lovaas therapy be supplied in the school setting (the school program did incorporate principles of applied behavioural analysis, but not pure Lovaas). The Hewko's request was not granted, and within approximately six months Darren began to regress. The school district recommended that Darren be removed from the integrated classroom and placed in a resource room at an alternate school, however, his parents refused and withdrew him from the public education system (Bakan, personal communication, Nov 22, 2005). The Hewko's sued the Province and the Abbotsford School District seeking relief for alleged discrimination pursuant to s. 15 and s. 7 of the **Charter of Rights and Freedoms** along with a claim of negligence and/or breach of duties provided for under the **School Act**.

**Reasons**

The Abbotsford School Board was found liable for breach of the Board's statutory duty "to consult with the plaintiffs" (para. 8) while the Court exonerated the Province from liability pursuant to statutory duties. In addition, the plaintiffs failed to establish a breach of either s. 15 or s. 7 of the **Charter**. In summary, the trial judge held that, "the District failed to carry out its duty to consult by failing to meaningfully consult with the parents" (para. 11) pursuant to the **School Act**.
Legal Reviews of Auton

Greschner & Lewis (2003)

Greschner and Lewis' article was one of the most cited papers during the Auton Supreme Court of Canada appeal. Their report examined the claims made in the lower Courts' proceedings in addition to critiquing the various components of the judicial reasoning process. The authors' rationale for challenging the outcome of Auton included: (a) its significant financial implications for society, (b) its role as setting precedent for future judicial involvement in policymaking, and (c) the need to find approaches for government that would make their policies incontestable under the Charter. Greschner & Lewis' review was based on the propositions that governments are better suited to determine complex policy issues rather than the Courts; that health care decisions should be evidence-based; and, that Courts could be useful in drawing attention to problems with government programs and governmental decision making.

Greschner and Lewis's critique began with an examination of the case within the contexts of health care and the Canada Health Act (CHA). Since Auton was a scope of coverage claim, when viewed in this perspective, it was essentially challenging the constitutionality of the CHA itself. To begin with, Greschner and Lewis pointed out that the principles of the CHA were frequently misused in the Auton trial. For example, Allan J. spoke on the primary importance of the medicare legislation being the provision of universal health care however, in actuality, it was the principle of comprehensiveness that was being litigated. For example, if Lovaas were an insured health service (principle of comprehensiveness) then all qualified provincial residents that required this treatment would have access to it (principle of universality). The authors argued that the principle of comprehensiveness was never intended to cover every medically necessary treatment.

Next, Greschner and Lewis took exception to the lower Courts' disregard for the implications of ignoring the CHA criteria for determining insured health services (defined as hospital services, physician services and surgical-dental services). Lovaas therapy is neither administered by a physician nor provided within a hospital setting. As an analogy, they pointed out that, under this direction, Provinces that do not have a prescription drug program would in fact be in violation of the Charter (since prescription drugs are medically necessary).

Greschner and Lewis also argued that there was a lack of evidence before the Courts that supported the conclusion of Allan J. that Lovaas therapy was a health issue (as opposed to an educational or social service issue). They pointed out that the determination of which camp Lovaas fell was integral to the procedures subsequently taken for the constitutional analysis. They reasoned that, due to the "deeper pockets" of the health care system, "it would presumptively be better for plaintiffs to approach the s. 15 analysis as patients, not as pupils or poor people" (p. 519).
The quality and type of evidence admitted was also identified as a major concern. Affidavits provided by the petitioners' parents were used to support their claim of treatment efficacy. However, Greschner and Lewis indicated that this anecdotal evidence was questionable due to the fact that these particular children may have improved anyway (due to maturation), or they may have been an atypical representation of autism. The authors also took exception to the Courts discounting what they perceived to be relevant research on the effectiveness of Lovaas. Their argument was that although the Supreme Court of BC ruled that the Court could not specifically order Lovaas therapy (hence, the question of its effectiveness became moot), the Court did order generic EIBI therapy in its place without an evidential foundation of its general effectiveness rate. Then, at the level of the Court of Appeal, the effectiveness evidence on Lovaas became especially relevant since the remedy granted to the four child petitioners was for Lovaas therapy.

The authors also questioned the utility of employing the general acceptance principle in judicial reasoning; the fact that other Provinces had developed IBI programs did not necessarily mean that IBI was effective. Essentially, policy precedence had become a surrogate marker for an effectiveness (or cost-effectiveness) claim. Instead, Greschner and Lewis argued that the Courts should have requested evidence from the government that supported their reasoning behind their program development. They also noted the conspicuous absence of any arguments about the relative effectiveness of either Lovaas or EIBI as compared to other treatment options for autism.

And finally, Greschner and Lewis delved into a detailed account of issues pertaining to the constitutional analysis by the Courts. However, as the purpose of this thesis centres on evidentiary issues, the constitutional analysis will not be discussed, with the exception of noting that Greschner and Lewis were highly critical of the lack of attention given to the Oakes analysis of the s.1 claim.

**Selick (2003)**

In Selick’s article for the Fraser Forum (a series of reviews of public policy issues by the Fraser Institute), the author lashed back at the Court of Appeal for BC’s decision in Auton (2002) by declaring that the judges failed to consider all the ramifications of their decision. Specifically, Selick spoke to the Court’s negligence of not reviewing opportunity costs for such a ruling. She contended that by funding Lovaas the additional money required would either have to be redirected away from other programs or alternatively, taxes would have to be increased. By reallocating money, Selick maintained that other faceless citizens would then become the next discriminated group.

The author also challenged the Court’s rationalization of its *parens patriae* jurisdiction (as defined in a previous section, this is a duty bestowed on the Courts to protect the rights of those perceived to be disadvantaged, e.g., children). Selick contended that, in its duty to protect
and advantage children, the government would inadvertently harm other children when funds were reallocated and/or taxes raised; she raised the question of whether the Courts should consider those children as well.

**Baker & Bhabha (2004)**

The article by Baker and Bhabha focused on the definition of *medically necessary* health services along with the Canada Health Act’s (CHA’s) principle of *universality* within the context of the Ontario health care system. The authors initially pointed out that the term *medically necessary* had not been clearly defined in either the Canada Health Act or in other provincial health insurance legislation. Consequently, determining the scope of medically required services is a policy decision as opposed to a legally enforceable rule. Another important distinction made was in regard to what constituted *experimental treatment*. The authors cited exceptions to the list of insured services in the *Ontario Health Insurance Plan* (OHIP) if a treatment for a medical condition was generally accepted to be experimental within Ontario (even if it is medically necessary and administered by a physician in a hospital). In a case that dealt with cancer treatments, the Province’s Health Services Appeal and Review Board outlined the following criteria in order for a treatment to be accepted as non-experimental (p. 27):

1. The procedure is accepted as appropriate in a number of respected academic and clinical cancer centres.
2. It is supported by published peer reviewed articles in respected journals.
3. It is publicly funded or funded by private insurers.
4. It is merely an extension of pre-existing and widely accepted therapeutic modalities.

When Ontario Courts are faced with decisions to fund procedures that are not listed under OHIP, the following types of evidence are used to determine whether or not the treatment should be considered experimental (p. 27):

1. Academic research noting the positive effects of the therapy.
2. The extent of the therapy’s usage and the length of time it has been practiced.
3. If the practice enjoys limited use, whether this is due to the high cost of the therapy and not doubts about its effectiveness.
4. The number of clinical trials.

Baker and Bhabha then proceed to examine the CHA principle of *universality* stressing that although everyone is covered, not *everything* that is deemed medically necessary is insured. However, similar to Allan J. in *Auton*, the authors approached the failure of government to cover Lovaas therapy through the lens of *universality* as opposed to comprehensiveness stating, “a [health] plan which provides necessary services to some, but not to others . . . is not only flawed in respect of the CHA, but also in respect of the Charter” (p. 32). They failed to recognize that Lovaas was not an insured health benefit that was being accessed by other groups.
Syrett (2005)

This article compared and contrasted Auton with a similar case litigated within the English Court system (T. v. Special Needs Educational Tribunal and Wiltshire County Council [2002] ELR 704 [Administrative Court]). Specific to Auton, the author identified many of the same issues discussed in previous critiques (confusion over the CHA principles of universality and comprehensiveness; lack of attention given to the efficacy debate). Syrett pointed out that the lower Court's reliance on parental affidavits of Lovaas treatment efficacy was an example of the legal system inappropriately generalizing individual outcomes to an entire population. He proposed that the judgments of the lower Courts may have been influenced by parental pressure and that some observers, "would accordingly be less prepared to regard this as a legitimate basis for funding the treatment than a decision which met standards of scientific rationality" (p. 350). Although Syrett supported the outcome of the Supreme Court of Canada, he contended that the efficacy debate once again did not receive the attention it demanded. He challenged the Court's line of reasoning in determining the medical necessity of Lovaas based on the fact that treatment had only been recently funded in two other Provinces in the past year (and hence, determined that it was still in the emergent phase). Syrett stated, "a governmental decision to fund or refuse a particular treatment may be taken for a number of reasons which are unconnected to the latter's efficacy" (p. 350). He continued by saying that the lower Courts put too much weight on the evidence provided by those that had vested interests in the outcome of the trial (the parents), while at the level of the Supreme Court of Canada, the focus appeared to shift to the possible cost consequences of ruling for a constitutional entitlement to Lovaas therapy. And in conclusion, Syrett hypothesized that parents of autistic children would have accepted the ruling of the Supreme Court of Canada better if the case had been framed within a clinical and cost-effectiveness context as opposed to a comparison of services between Provinces (general acceptance principle).
Objectives

The objectives of this thesis are, (a) to examine the role of evidence within the contexts of law, health policy and health care, and (b) to determine the current state of knowledge in regard to the effectiveness claims of Lovaas therapy for autism spectrum disorder.

Rationale

The Auton case drew attention to the contentious interplay between law, health policy, and health care in regard to the Courts’ involvement in health policy decisions. This played out through the process of judicial policymaking. While originally called upon to decide cases of medical malpractice, the Court’s role has evolved to include issues regarding the regulation of medical practice, physician supply management, hospital restructuring, and access to specific treatments (as was the case in Auton). Proponents of judicial policymaking suggest that the process protects minority rights, the promotion of humane conditions in institutions, restricts bureaucratic arbitrariness and in general, promotes positive social change (Anderson, 1992). On the other hand, critics cite limitations inherent in the legal system such as: (a) judges’ and lawyers’ lack of educational background or expertise to handle complex policy issues; (b) the nature of the adversarial system whereby information can be withheld; (c) the Courts’ reliance on precedent cases to decide future proceedings; (d) the limited focus of a judicial review process; and (e) the Courts’ inability to foresee the long-term consequences of their decisions (Anderson, 1992). Manfredi and Maioni (2002) concur with Anderson in regard to the adversarial structure of adjudication impeding comprehensive information gathering. However, they also contend that within the context of rights discourse, judicial review narrows the range of policy alternatives and systematically favours national norms and standards.

In law, adjudicative facts provide the basis of determining historical cause-and-effect relationships; however, many academics contend that Courts are ill equipped to extrapolate the process to present day phenomena. For example, in Auton, the systematic review and critical appraisal of the evidence concluded that there was a lack of scientific rigor to support the effectiveness claim for LAT (Bassett et al, 2000). However, despite these findings, five judges (two trial judges and three appeal judges) proceeded to award Lovaas treatment to the petitioners. However, of particular interest to this thesis is the reversal of the lower Courts’ decisions by the Supreme Court of Canada. How is it the same evidence produced such a different outcome? What was going on in the lower Courts that yielded findings of fact (formal legal truth) that deviated from the actual truth (substantive truth)? These are important questions as the ramifications of a judicial policymaking process can extend far beyond the original case at hand by redirecting public money to specific stakeholders.

One of the key issues in a judicial review is that the domains of law, policy and health care differ in regard to their respective world views, conceptualizations, interpretations and use
of evidence. This phenomenon has been the focus of debate (Eisenberg, 2001; Havighurst et al., 2001) and gained momentum as Auton advanced to the Supreme Court of Canada.

This thesis was specifically designed to investigate the conceptualizations and processes used by each of the domains (law, health policy and health care) in order to gain an understanding of how each seeks, understands and applies evidence. This is accomplished in two parts. First, the legal dimension of the health technology assessment (HTA) framework utilizes a qualitative grounded theory methodology to examine participant interviews and legal documents. This provides early conceptualizations of each sector’s perspectives on evidence along with providing HTA researchers with an approach to investigate the social (legal) context. And second, the effectiveness dimension of the HTA framework employs the methods of systematic review and critical appraisal to investigate the ongoing Lovaas Autism Treatment effectiveness debate. Together, these two components provide insight into how the Lovaas effectiveness evidence was handled in the Auton case, and how each sector conceptualizes the evidentiary process in general.

The goals of this research are:

1. To be able to provide policymakers with a current, comprehensive body of evidence on Lovaas Autism Treatment to aid in the development and evaluation of provincial EIBI treatment programs.

2. To be able to provide lawyers with a current, comprehensive body of evidence on Lovaas Autism Treatment for potential future litigation.

3. To create a better understanding within and between the domains of law, health policy and health care in regard to how each conceptualization seeks, understands and applies evidence.

**Research Questions**

This thesis sets out to answer the following questions:

1. How do the sectors of law, health policy and health care seek, understand, and apply evidence?

2. What is the current state of knowledge in regard to the effectiveness claims of Lovaas therapy?

Chapter 1 has provided an introduction to Autism Spectrum Disorder and Early Intensive Behavioural Intervention. It examined one specific form of applied behavioural analysis – Lovaas Autism Treatment – and highlighted the ongoing effectiveness debate. The chapter proceeded to situate this intervention within the local health policy context of the British Columbian autism intervention program along with the legal context of the Charter challenge to access this therapy (Auton et al.).
Chapter 2 goes on to identify and summarize the relevant multi-disciplinary conceptual perspectives on evidence uncovered in the literature along with the opposing conceptualizations that accompany each view. Processes for seeking, understanding and applying evidence in each of the legal, health policy and health care domains are discussed. From this literature review, preliminary conceptual maps and frameworks are developed to provide an analytical lens to examine the role of evidence within the legal context of the Health Technology Assessment framework.

Chapter 3 provides an in-depth description of the research methodologies utilized in this thesis: (a) grounded theory, (b) systematic review, and (c) critical appraisal. It also provides background information on the qualitative research software program – QSR N6 – that was used to assist in the analysis of the interview data and legal documents. In sum, this chapter outlines the processes and rationale behind the design, conduct and analysis of this thesis.

Chapter 4 presents the results of the analyses of both the interviews and the legal documents, along with the outcomes of the critical appraisal of the Lovaas evidence. The qualitative data findings from the interviews and legal documents are framed within the context of the conceptual maps developed for this thesis. The findings from the systematic review and critical appraisal of the Lovaas effectiveness evidence are applied to objective criteria to discern quality and validity.

And finally, Chapter 5 provides a discussion and conclusion to the analyses of the role played by evidence in legal, health policy and health care contexts along with the findings from the critical appraisal of the effectiveness evidence on Lovaas Autism Treatment. In addition, this chapter highlights how legal issues are identified and analyzed within a health technology assessment framework. The chapter also draws conclusions on the current state of knowledge and strength of the body of evidence in regard to Lovaas Autism Treatment. And finally, this chapter concludes by proffering suggestions for further research.
CHAPTER 2: CONCEPTUAL FRAMEWORK

Understanding the Role of Evidence in Law, Health Policy and Health Care

Increasingly, law is relying on scientific evidence to inform its fact-finding process. This has occurred due to the technology explosion, adoption of standardized rules of evidence, and the increasingly more common litigious approach to problem resolution. While intuitively inviting science into the evidential forum seems like a natural course, the process has not been without its problems and controversies. As described by Kaye (1997), law has "harnessed the powerful engine of scientific investigation to procedures crafted in the days of the horse-drawn cart" (p. xvii).

In order to examine the role evidence played within the three contexts of law, health policy and health care in the Auton and Anderson legal proceedings, I developed a conceptual framework that pulled from relevant concepts within each of these three domains. These concepts were selected based on their ability to provide insight into the processes used by each sector to seek, understand and apply evidence. This chapter describes these concepts, presents the current discourse on each topic, and then proffers a conceptual framework utilized for the analysis in this thesis.

The Concept of Evidence

The concept of evidence has different meanings and standards depending upon which disciplinary camp one is from. From the health care perspective evidence is based on empirical observations of real events (Eddy, 2001). Ideally this plays out in the form of randomized controlled trials. However, when this level of evidence-based data are not available, alternate processes such as peer review and consensus development are employed to gather, assess and disseminate other types of evidence (to be discussed in a later section). However, the utility of each strategy comes with its own unique strengths and limitations.

In the policy context, the concept of evidence has been classified as either colloquial or scientific (Lomas, Culver, McCutcheon, McAuley & Law, 2005). Colloquial evidence is the non-scientific contextual side of a policy, "anything that establishes a fact or gives reason for believing something" (p. 3). Alternatively, scientific evidence is seen as, "information generated through a prescribed set of processes and procedures recognized as scientific" (p. 8); it is knowledge that is explicit (codified and propositional); systematic (utilizes transparent and explicit methods); and replicable (utilizing the same methods on the same sample will produce the same results). This systematic review identified differences between these two types of evidence; the first was in regard to its usage. The authors described colloquial evidence as being incorporated into policy decisions once decision-makers agreed upon its relevance, while the inclusion of scientific evidence was methodologically determined. The authors also identified two different views pertaining to the role of scientific evidence. The first perspective, which has
its genesis within evidence-based medicine, was that science reveals "universal truths" (p. 10) based on the assumption that its role is independent of context (context-free evidence). This view addresses the question "can it work?" and "provides a glimpse of what might be achieved under ideal circumstances" (p. 10). The second point of view proffered by the researchers originated from the social sciences and holds as its conviction the belief that, "evidence has little meaning or importance for decision making unless it is adapted to the circumstances of its application" (p. 11). This type of evidence seeks to answer the questions "will it work?" and "is it worth it?" and is termed context-sensitive evidence. The authors concluded that:

To create context-sensitive guidance [to policy-makers], context-free science needs to be integrated into the science on "local" variables such as public attitudes, patient preferences, professional proclivity, managerial capacity, economic feasibility, geographic location, and so on, in search for consensus around what might be achievable rather than what might be a universal clinical truth. (p. 11)

Lomas et al. (2005) also summarized the different categories or dimensions of evidence uncovered in the literature. These consisted of three general approaches to defining evidence: (a) by method of collection, (b) by general-purpose, and (c) by source. Method of collection includes evidence compiled through research designs such as experimental or observational; the general-purpose approach takes into consideration the reason for its application, such as problem identification as opposed to measuring effectiveness; while, source differentiates research from colloquial forms of evidence.

The authors go on to describe six dimensions of context to be considered alongside the core concepts of health outcomes and appropriateness. These contextual dimensions include: (a) implementation evidence, (b) organizational evidence, (c) attitudinal evidence, (d) forecast evidence, (e) economic/financial evidence, and (f) ethics evidence. While the core concepts are obtained through experimental, quasi-experimental, or observational methods, their literature review supported the argument that existing research methods could be easily employed to obtain additional information on each of the six contextual categories.

And finally, understandably, evidence is the most conceptually developed within the discipline of law. Evidence is broadly defined as, "all the means by which any alleged matter of fact, the truth of which is submitted to investigation at judicial trial, is established or disproved" (Gifis, 2003, p.182). Evidence can be defined by its source (expert evidence); type (hearsay evidence); degree of proof (conclusive evidence); process for obtaining (illegally obtained evidence); quality (incompetent evidence); and weight (preponderance of evidence), to name but a few. Law schools offer numerous courses dedicated strictly to this concept while legal libraries are stacked with authoritative texts on the subject. However, some policy-makers contend that legal definitions of evidence were not very helpful when it comes to guiding policy decisions (Lomas, et al., 2005).
Opposing Conceptualizations of Evidence

It was inevitable that when the domains of policy and health care intersected within the judicial policy making context of the Auton and Anderson trials that each sector's differences in regard to their conceptualization of evidence would become apparent. Policy's confidence in their commissioned health technology assessment of the evidence on Lovaas was usurped by the apparent weight placed upon the parents' affidavits by the Courts. Legal rules of evidence, tools for constitutional analysis, and concepts such as the burden of proof contrasted sharply against processes employed by the policy and health care sectors to support their evidence in the Lovaas efficacy debate.

Conceptual Background for Thesis

In April 2000, the American Institute of Medicine (IOM) and the Agency for Healthcare Research and Quality (AHRQ) spearheaded a debate called "Evidence: Its Meanings in Health Care and in Law" (Havighurst, et al., 2001; Peterson, 2001). This forum brought together leading experts in the fields of law, epidemiology, health services research, and health plan administration to explore the differing professional constructs of evidence. The impetus for the workshop was the paradigm shift in health care towards evidence-based medicine and the outcome of a precedent U.S. Supreme Court case that saw judges assuming a greater role in screening admissible expert evidence (Daubert v. Merrell Dow Pharmaceuticals, 1993). The workshop participants shared a common concern that, "legal uses and interpretations of science-based medical evidence . . . may diverge substantially from the uses and interpretation of that evidence by the medical and health care researchers who produce it and of the practitioners and health plans that use it in making clinical decisions and policies" (Havighurst, et al., 2001, p. 195).

Through a series of reports commissioned before the workshop, in addition to academic debates throughout the sessions, the IOM/AHRQ participants discovered that opposing constructs of evidence existed amongst various professional bodies. Although the participants were unable to reconcile their differences, six specific conceptual areas of discourse were identified: (a) population probabilities vs. individual causation, (b) pre hoc vs. post hoc evidence, (c) clinical progress vs. legal reform, (d) adjudication of differences, (e) rules of evidence; and, (f) decision-maker. The workshop concluded, "open-minded discussions and thoughtful conceptual and empirical analyses can bridge our fields and identify areas of miscommunication." (Eisenberg, 2001, p. 380). The commissioned reports, along with the findings from this workshop, were published collectively as a series of peer-reviewed papers in a special feature in Journal of Health Politics, Policy and Law (Vol. 26, No. [2]). For ease of in-text referencing, this body of knowledge will be cited as Eisenberg (2001).
Population Probabilities vs. Individual Causation

To begin with, the IOM/AHRQ forum determined that law, policy and health care differed in their conceptualization of the patient/client by taking either a population or individual perspective. In tort law (negligence litigation), the Court seeks to determine whether harm has come to an individual by way of an action that has already occurred, or whether the individual was denied a potential benefit from an action that was not carried out. Eisenberg (2001) coined this viewpoint, “evidence of the instance” (p. 375) as it focuses on the tenets of individual rights, wrongs, and harms and utilizes evidence in an effort to determine causation under particular circumstances. (However, in Auton the lower Courts were assuming a policy role, as opposed to the familiar “individual causation” perspective).

An example of law’s individualist perspective can be found in an American case in which expert evidence pertaining to the use of facilitated communication (FC) to question an autistic child in Court was deemed inadmissible (Candelora, 1995). (“Facilitated communication involves hand-over-hand or hand-on-forearm support of students by a facilitator . . . to offer physical support while typing on a keyboard” [p. 753]). Because the Court viewed FC as scientific, it was subject to the same admissibility criterion of reliability. Critics felt that, “by requiring general quantitative testing [of FC], the Courts . . . shifted the focus from an individual’s [italics added] ability to communicate to the ability of a class of people to communicate” (p. 99).

Within the policy forum, decision-makers are compelled to discern the fairest way of equitably distributing finite resources amongst a population of individuals. Within this population perspective, many competing needs must be weighed through various decision making processes.

And, in regard to health care, in response to the evidenced-based medicine movement, physicians are increasingly relying on population-level statistics based on probabilities in order to make clinical decisions; this equates to “evidence of the generalizable” (Eisenberg, 2001, p. 375).

Pre Hoc vs. Post Hoc Evidence

Another differing perspective that arose was how the different disciplines conceptualized evidence in regard to the timing of utilization of evidence relative to the outcome. In tort law, evidence is used post-hoc to discern whether an outcome that has occurred in the past was caused by an appropriate, harmful or inappropriate action. It is used to “judge responsibility and render justice” (Eisenberg, 2001, p. 376). While in Auton, evidence involved pre hoc assessments by the Courts. In policy forums, evidence can be used pre hoc to inform decision-makers on population-level issues prior to their outcomes. In health care, evidence is also generally used in a pre hoc probabilistic fashion to support a clinical decision.
prior to a patient outcome (however, Eisenberg noted that medicine often reflects back on past decisions to learn from its mistakes [e.g., morbidity and mortality rounds]).

**Clinical Progress vs. Legal Reform**

The way in which change occurs in law and health care also impacts the conceptualization of evidence. In law, Courts, legislatures, or regulators perform the change agent role, while in health care, change may occur when new scientific evidence is adopted by key opinion leaders and disseminated to other practitioners. Within the policy context, change occurs through processes of policy debate.

**Adjudication of Differences**

Another identified difference was the way in which experts were used in each profession to adjudicate differences. The judicial system operates within two different sets of ideals about the conduct of fact-finding proceedings. The first is the traditional adversarial system and the second is the gatekeeper approach (Shuman, 2001). These two processes represent different methods to address the core values inherent in the judicial system: accuracy, fairness, efficiency, consistency, and accessibility. Which approach is used often depends on the type of case (e.g., criminal vs. civil) and the evidence at issue. An understanding of these two different perspectives is important in order to discern how law assesses expertise and how evidence is admitted into Court.

In the adversarial approach, lawyers decide on what evidence to submit and judges decide which evidence can be presented to and ultimately admitted by the Court. The adversarial process assigns responsibility to the individual parties (and not the judge) for identifying relevant issues, presenting proof, and championing the case through the judicial system (Brooks, 1991). Each party is responsible for finding and submitting evidence in addition to finding limitations or weaknesses in the opposition's case. The underlying assumption is that truth is more likely to be uncovered by a contest between those that hold the greatest vested interest in the outcome. Critics of the adversarial system argue that both lay and expert witnesses called by each respective party will favour their attorney-employer's perspective. Historically, admissibility of expert opinion was based on an assessment of the expert's qualifications leaving an evaluation of the reliability and weight of their presented evidence up to the trier of fact. However, because the individual parties control the evidence, there is a risk that the trier of fact may be exposed to a biased sample of experts. It has been proposed that one alternative for civil cases would be Court-appointed experts, a process similar to a journal editor requesting experts to review a manuscript. In theory, this would provide impartiality with judgments based on expertise rather than advocacy (Eisenberg, 2001). However, this suggestion has been met with much controversy and as a result, is rarely used. In addition, judges may also be unable or unwilling to weed out poor experts, as there exists a large pool of "professional experts" more than willing to make a career out of testifying.
The gatekeeper model arose from the outcome of three precedent American cases: *Daubert v. Merrell Dow Pharmaceuticals* (1993); *General Electric Co. v. Joiner* (1997); and, *Kumho Tire Co. v. Carmichael* (1999); within a Canadian context, the model is rooted in *R. v. Mohan* (1994). Although this approach to adjudication will be described in more detail in a subsequent section, suffice it to say at this point that the outcome of these cases resulted in judges assuming a far greater role in ensuring that evidence presented to the Court was reliable and relevant (Eisenberg, 2001).

In the health care context, scientific evidence is generally integrated once it has been through a peer review process; one such method is publication in a peer-reviewed journal. Here, potential manuscripts are distributed amongst substantive and methodological experts for critique with the final decision on their worthiness for publication residing with the journal's editor. In this scenario editors become the adjudicator of evidence for their profession.

The IOM/AHRQ workshop did not report on the concept of adjudicator for the policy context, however, this writer is hypothesizing this role to be assumed by bureaucrats and looks to discover this in the data.

**Rules of Evidence**

One of the main focuses of the IOM/AHRQ conference was the changes to evidentiary rules brought about by the three precedent American cases. Although *Daubert* contributed to the codification of the rules of evidence, subsequent cases such as *Kumho Tire Co. v. Carmichael* demonstrated that judges had flexibility in applying the test of reliability to the evidence and substantial discretion in determining which evidence to admit. Despite great expectations of raising the evidentiary bar for admissibility of medical evidence, the impact of *Daubert* (and *Mohan* in the Canadian context) was not consistent across the Courts. In fact, legal rules of evidence are in constant flux.

Within health care, rules of evidence focus on the critical appraisal of the available scientific evidence. This can occur in a couple of different ways. At the level of the individual physician, this would involve an evaluation of a particular article in respect to its relevance to a patient's specific clinical problem. Ideally, the article would be a systematic review of a number of well-designed randomized controlled trials (RCTs), which also incorporated an assessment of the level or quality of evidence. A review of the level of evidence is a process that allows for the qualitative categorization of research evidence in addition to defining specific grades of recommendation, while a review of a study’s quality reveals the measures researchers have taken in their design, conduct, and analysis in order to limit bias (to be discussed in more detail in the Methodology chapter of this research).

It is recognized that this systematic process of appraising research does not always occur in actual medical practice. Some medical specialties, such as surgery, may not be conducive to evaluation by RCTs (unless the principle of equipoise exists between two surgical procedures).
Other reasons for the difficulties in designing surgical RCTs include selection and observer bias, blinding, learning curve, effectiveness versus efficacy and standardization of technique (Meakins, 2002). Thus, in these situations a weaker study design (such as an observational study) may be used. However, even for research questions that are conducive to being examined with an RCT design, factors such as costs, length of follow-up, and ethical issues surrounding randomization, may often prevent it.

Although peer-reviewed evidence in medicine is held in high regard, there does not prevail common criteria ("rules of evidence") upon which to evaluate its quality and strength. Instead, multiple frameworks exist upon which clinical practice guidelines are often developed. One example is the methodology developed by the Canadian Task Force on the Periodic Health Examination in 1979 for determining strength of evidence and levels of recommendation (1994). This methodology was subsequently adapted by the U.S. Preventive Services Task Force (Harris, et al., 2001) and since its inception the two countries have collaborated on their recommendations.

The IOM/AHRQ workshop did not report on the concept of rules of evidence within a policy context.

**Decision-maker**

While both law and medicine value the credentials of experts, some differences exist in regard to the source of the evidence and who decides on its validity; essentially, who should determine which evidence should be followed? In law, the source of evidence is often times a primary source (e.g., the author of a research paper, or a credible expert in the field). However, the evaluation of the strength of the evidence can rest on the shoulders of a less well-educated jury (e.g., civil trials in BC) or on a judge's assessment of the credibility of an expert witness or their evaluation of the scientific merit of a study. Determining credibility of health care practitioners can be especially difficult in that, beyond their generic license to practice, there is no formal mechanism to evaluate their qualifications and credentials. When called upon as an expert witness, the practitioner's assertion of their qualifications is taken at face value ("prima facie") (Supreme Court of BC, Aug 30, 1993).

In medicine, those who are deemed the most expert judge the evidence. Physicians infrequently will read the primary source of the evidence (e.g., the research paper); instead, relying on protocols, decision trees, consensus statements, clinical pathways and clinical practice guidelines. Physicians also utilize secondary reviews and syntheses of research, such as Cochrane reports (The Cochrane Collaboration is an international, non-profit organization that produces and disseminates systematic reviews of health care interventions).

And, once again, the IOM/AHRQ workshop did not specifically identify the decision-maker in health policy but is assumed to be the bureaucrat and/or elected official.
Definitions of Effectiveness

Another differing conceptualization that did not originate from the IOM/AHRQ workshop, but instead became apparent through this literature review, was the conflicting definitions of effectiveness. Members of the health care team would view a treatment as effective if it matched the specific requirements of an individual patient, while health policy researchers would view effectiveness in light of increasing health utility indicators as demonstrated through randomized controlled trials (Biller-Andorno, Lie, & ter Meulen, 2002). In addition, these authors also extend this concept to include differing conceptualizations of cost effectiveness arguing that recommendations produced from cost effectiveness analyses may conflict with physicians’ pursuit of optimal individual efficiency. These two views also link with the previous population probabilities vs. individual causation debate.
Processes of Seeking, Understanding and Applying Evidence: Legal Context

Judicial Policy Making

Auton took place under the jurisdictional context of judicial policy making. This is a process whereby the Courts ascertain legal jurisdiction over deciding a public policy issue with the purpose of producing socially desirable results (Feeley & Rubin, 1998). While not exclusive to rights-based judicial review brought about by Charter challenges, within the realm of health care these cases have contested issues such as regulation of medical practice, physician supply management, hospital restructuring, and access to specific treatments. It is within this context of judicial policy making that the legal processes for seeking, understanding and applying evidence will be examined.

Manfredi and Maioni (2002) postulated that the judicial policy making process raises three specific issues: (a) rights discourse narrows the range of policy alternatives, (b) the adversarial structure of adjudication impedes comprehensive information gathering, and (c) rights-based judicial review systematically favours national norms and standards. Of particular interest to this research is issue b.

Manfredi and Maioni (2002) proposed that the institutional attributes of adjudication impact judicial policy making by impeding comprehensive information gathering and processing. Judicial process is limited to a retrospective examination of historical/adjudicative facts pertaining to the issue at question rather than a comprehensive social/legislative perspective required for policy making. Adjudicative facts provide the basis of determining historical cause-and-effect relationships. Manfredi and Maioni contended that "a tension [exists] between the type of analysis needed to solve complex and multifaceted social problems and the techniques used in the judicial process to gather, process, and evaluate information" (p. 216). In particular, one attribute of adjudication - incrementalism- can work against the system. In typical litigation, the Courts contribute to the advancement of legal-moral principles incrementally through case-by-case analysis and judgment. This process allows Courts to evaluate the outcomes of previous cases before applying the same principles to the next case thus avoiding large-scale policy decisions. However, with judicial policy making, the relief sought is usually widely applied, so policy decisions have the potential for a more significant impact.

Manfredi and Maioni (2002) also postulated that cases brought before the Courts for rights-based judicial review are not necessarily representative of the population in question. In addition, they surmised that judicial policy making is plagued with poor compliance by the involved parties and institutions due to the difficulty the current legal system has in gathering and processing the myriad of social/legislative facts required for policy making. And finally, they also pointed out that the Courts are unable to initiate policy review to determine the unintended consequences of previous Court-mandated policy decisions.
Anderson (1992) contributed to the discourse on Court-directed social policy. This author contended that not only can important information/evidence be filtered or withheld from the judicial system, but also, the Courts may rely too heavily on academic theoretical arguments as opposed to listening to practitioners in the field who would be more cognizant of the "real world" scenario and the inherent limitations of social science theory. Decentralization of the legal system is another concern that has some wondering how deciding individual cases impacts society as a whole; disjointed policies and different treatment of similar cases are but a couple of examples.

**Legal Discovery**

One method lawyers use to seek evidence in the pre-trial stage of a case is through discovery, which is: "the formal and informal exchange of information between sides in a lawsuit" (Oran, 1991, p. 88). Discovery is a procedural stage used in civil proceedings whereby parties must disclose all relevant facts, documents and other specifically described evidence pertaining to a case. It is also used in criminal cases where the Crown is required to provide pre-trial discovery by either informal production of evidence obtained during the criminal investigation or at a preliminary hearing.

Discovery is comprised of three basic forms: written discovery, document production, and examination for discovery (called oral depositions within the American context).

**Written discovery** can include Interrogatories and Requests For Admissions. Interrogatories are a pre-trial discovery tool whereby a party prepares written questions and serves these questions on the other party to be answered under oath or solemn affirmation. Request For Admission is another device used at the pre-trial stage of the proceedings to ask a party for either a positive affirmation or a denial of a material fact or allegation at issue.

**Document production** is the right of both parties to have access to any document in the possession or control of the other party that may in any way be relevant to the issues raised in the case. More recently, Courts have required that parties also produce electronic documents and files including emails; this has subsequently been termed e-discovery.

In Canada, an examination for discovery, in the absence of a specific Court application, is limited to the parties to the proceeding, while in the American legal system, depositions may be taken from not only the parties to the action but also from any witness who may have evidence that is relevant to the issues in the case. In both countries, examinations for discovery/depositions require, "a statement of a witness under oath, taken in question and answer form as [evidence] would be [presented] in Court, with opportunity given to the adversary to be present and cross-examine" (Gifis, 2003, p. 141). Often civil and criminal cases end up being resolved after discovery as each side becomes aware of the strength of their opponent's case.
Rules of Evidence

Legal rules of evidence originate from statutes and precedent Court cases. On the American front, historically, witnesses were initially only allowed to testify about what they perceived through their five senses. In addition, expert witnesses had to be qualified by training, knowledge, skill, or experience in a particular subject. The rules of expert evidence in the United States subsequently changed in 1923 with the outcome of *United States v. Frye*. In this case the defendant attempted to introduce a new technology into evidence (the systolic blood pressure deception test). Since the Court had never heard of this test, it ruled that any scientific principle introduced into Court had to demonstrate general acceptance by the scientific community (the "general acceptance standard") (Lee, 1997). General acceptance required researchers to quantitatively prove a technique's reliability. However, for new procedures that had not yet achieved the level of general acceptance, some Courts could apply a relevancy standard instead. While this process still required a proof of reliability of the evidence, to substitute for its lack of general acceptance, the Courts would look at the novelty of the procedure, the literature on its merits, the qualifications of the expert witness, and the non-judicial uses of the scientific technique (Candelora, 1995).

The general acceptance rule of *Frye* continued to be followed by the majority of American Courts despite the adoption of the *Federal Rules of Evidence* in 1975. The goals of the *Federal Rules* were, “to secure fairness in administration, elimination of unjustifiable expense and delay, and promotion of growth and development of the law of evidence to the end that the truth may be ascertained and proceedings justly determined” (Rule 102). These legislatively enacted rules, which pertained only to federal trials, greatly relaxed the rules surrounding admission of expert scientific testimony by removing the general acceptance condition. In its place, Rule 402 introduced the concept of “relevancy” of evidence. In addition, Rule 702 governing expert testimony also made no reference to general acceptance as the absolute prerequisite to admissibility. The *Rules* also dictated that it was the responsibility of the judge to determine whether or not admissible evidence was relevant and reliable.

The Frye test was applied within the autism context in a couple of American cases that also challenged the admissibility of evidence on facilitated communication (FC) (Candelora, 1995). The Courts ruled that the FC technique was scientific in nature and therefore subject to the Frye test. This required proof of the technique’s reliability and general acceptability. However, in a similar case the following year, the Court decided that FC was not based on scientific techniques, therefore not bound by the Frye test of admissibility. Instead the judge ruled that FC was a special mode of communication (similar to sign language). In this case, the Court focused on issues such as whether the witness was competent to testify and whether the facilitator was qualified to administer the technique.
In 1993, the American legal tradition was once again challenged in the Supreme Court with *Daubert v. Merrell Dow Pharmaceutical* (a case about birth defects allegedly caused by a prescription medication). In this litigation, the lower Courts had excluded evidence based on its failure to comply with the *Frye* criteria for admissibility. However, the Supreme Court ruled that in actual fact, the *Federal Rules of Evidence*, as opposed to *Frye*, should have guided the evaluation of the expert scientific testimony. This ruling subsequently negated the general acceptance rule in all Federal Courts and in any state that also followed this standard (Gorman, 1999). Judges now had greater freedom to assess evidence than was previously afforded under *Frye*. However, the onus for ensuring the scientific merit of evidence presented by expert witnesses now resided with the trial judge. To aid this process, the Supreme Court established four criteria for determining whether expert testimony was indeed "scientific knowledge": (a) whether the theory or technique can be (has been) tested, (b) whether the theory or technique had been through the peer review and publication processes, (c) whether the known or potential rate of error had been established for a particular technique, and (d) whether the evidence had been accepted by the scientific community (Bruce, 1999). The actual weight assigned to each criterion, however, was left up to the Courts to determine. The Supreme Court's opinion was that federal judges possessed the capacity to critically examine an expert's evidence in order to determine its reliability and relevancy (Eisenberg, 2001).

Essentially, the outcome of *Daubert* reformed the American legal system from an adversarial to a gatekeeper model with its purpose to shield jurors from junk evidence. This rule of evidence became known as the "scientific soundness standard" (Kaye, 1997, p. 87), and was further supported by two more American cases: (a) *General Electric Co v. Joiner*, and (b) *Kumho Tire Co. v. Carmichael*.

The *Daubert* decision stood boldly in the face of seventy years of precedence in the adversarial system where general acceptance by the respective community had been the only criterion for assessing the quality of evidence. While it has resulted in a higher threshold for the admissibility of medical expertise in civil litigation in toxic tort and products liability cases, its application has not been consistent across the American legal system. For example, it has had a minimal effect on medical malpractice cases. Shuman (2001) theorized that adoption was slow due to pragmatic reasons, such as a Court's inability to wait for a sufficient body of evidence to be developed before deciding a case, and due to differing philosophical viewpoints. In fact, Shuman made the analogy between *Daubert* and science-based medical evidence; while both claim to be evidence-based neither has been widely embraced.

*Daubert* has been challenged on the grounds that the skill set required to carry out an effective evaluation of the evidence is not part of the current legal curriculum. In addition, from a mere practical viewpoint, the additional requirement of critical review would put more
pressure on an already over-loaded judiciary system. In all, Daubert brought about the following judicial changes to the American Courts (Lee, 1997):

1. Judges now had to determine whether a jury could hear expert testimony, however, they did not need to decide causation.
2. Expert witnesses were now restricted to only testifying about scientific evidence, thereby making it difficult for plaintiffs to "get to the jury".
3. Some judges now required additional training in the scientific process.

Before American Courts apply the Frye test, relevancy test, or the Federal Rules of Evidence, they first classify the evidence as either "hard" or "soft" (Candelora, 1995). Hard evidence, such as physiological or genetic studies, is seen as producing "conclusive, objective results ... quantifiably based on nonhuman, objective testing devices" (p. 100); while soft evidence is viewed as opinion evidence based on the study of human behaviour, which is "subjective and incapable of scientific verification" (p.: 100). Hard evidence is amenable to quantitative testing while soft evidence is studied qualitatively to discern its general accuracy. Classification of evidence assists the Courts in applying the correct test to determine its admissibility.

R. v. Mohan (Supreme Court of Canada, 1994) is the Canadian equivalent to Daubert. In this case (which centered around sexual assault charges laid against a pediatrician) the lower Courts held that expert opinion proffered by a psychiatrist in regard to a psychological profile of the accused was inadmissible. On appeal to the Supreme Court of Canada, the Court addressed the issue of admissibility of expert evidence through an examination of the rules pertaining to both expert evidence and character evidence. Specific to expert evidence, Sopinka J. stated that admission of evidence was dependant upon the application of four criteria: (a) relevance, (b) necessity in assisting the trier of fact, (c) absence of any exclusionary rule, and (d) a properly qualified expert (p. 2). The concept of relevance was defined as "so related to a fact in issue that it tends to establish it" (para. 18). However, logically relevant evidence could be excluded if it involved too much time relative to its value, or if it could be misleading to the trier of fact. Relevance was further defined to include a test to determine whether it was essential and reliable. Expert evidence now had to contain technical knowledge that would make it difficult for a judge or jury to reach a conclusion without its assistance; and, then, the evidence had to be evaluated further to determine its reliability. Although Sopinka J.'s interpretation of the characteristics of expert testimony paralleled that of the U.S. Federal Rules of Evidence, it left many key concepts such as "special knowledge", "reliability", "novel scientific theory" and "technical matters" open to interpretation by the lower Courts (Bruce, 1999). One example of the application of Mohan can be found in R. v. D.D. (Supreme Court of Canada, 2000) whereby the Court narrowly held that evidence presented by a psychologist in a child abuse case was neither relevant nor necessary as the content of the testimony itself had no technical quality to it.
And finally, another relevant rule of evidence pertained to judicial notice. This doctrine rules that the Courts can, "admit as 'proved' such facts that are common knowledge to a judicial professional or to an average, well-informed citizen" (Gifis, 2003, p. 277). The proffered example was that the Court could take judicial notice that mail is not delivered on New Year's Day. Therefore, "a judge [can] accept the existence of a proposition of fact even though no party has proven it by admissible evidence" (Sheppard, 1996, p. 783).

**Burden of Proof**

Within a legal context, burden of proof is based on the most probable scenario that will "reduce the risk of an incorrect judgment in the direction in which it is considered worse to make a mistake" (Eng, 1999, p. 487). For example, it would be viewed as a "better" decision to let ten guilty criminals go free than to mistakenly imprison one innocent person. The burden of proof rests with the party that is demanding judgment with the more unfavourable consequence.

Some legal rules covering the burden of proof criterion stipulate that, when faced with competing descriptive characterizations of an event, the one with the most probability must be used as a basis. This standard of proof is referred to as a "balance of probabilities", or a "preponderance of evidence/probability" (e.g., 50% plus at least 1%). However, at other times rules may be more stringent, thus requiring more than a simple preponderance of probability (e.g., "clear and convincing" evidence); these rules are usually reserved for the more serious cases.

Determining "truth" in a Court of law would appear to be an obvious objective of the judicial system. However, the concept of truth is elusive. In law there exists two definitions of truth: (a) substantive truth (the actual truth), and (b) formal legal truth (whatever is determined to be factual by the legal fact-finder, whether or not it supports the substantive truth) (Summers, 1999). Therefore, once a Court rules on matters of fact, these facts (or truths) now exist in the legal sense (formal legal truth) although they may not exist in reality (substantive truth). Ideally, the goal is to have congruency between substantive truth and formal legal truth, but this is not always the case for a couple of reasons. First, substantive truth and formal legal truth may diverge not by design but due to the particular circumstances of a trial. Some examples of intervening variables would include unequal lawyer representation, unequal financial resources for trial preparation, prejudice or biases of fact-finders, and the unavailability of evidence. All these factors could impact the final outcome of a trial.

Second, substantive truth and formal legal truth may diverge due to actual trial Court procedures and rules of evidence. For example, in criminal cases, if evidence were obtained illegally it would be inadmissible (although it may be relevant to determining the facts of the case). Therefore, due to the exclusion of this evidence, a Court may have difficulty determining the "truth". Adjusting the facts out of a desire for concrete reasonableness and justice is another example of a trial Court procedure that could impact fact-finding. Here, the person applying
the law decides to take as a basis another descriptive characterization of the phenomenon under question (rather than the most probable one) in order to avoid a result that they would consider to be unreasonable or unjust (Eng, 1999). This assessment involves balancing what the probability is for a particular descriptive characterization against what probability will be required for the characterization to be usable as a basis (burden of proof). The reasoning being that the quest for just results often influences either the assessment of evidence or the determination of burden of proof. Balancing probabilities helps determine which option would produce the most harmful effects if a mistake were made and then this option is avoided. This results in a decision that may not necessarily be based on the most probable scenario.

**Standard of Care: Expert Witnesses and Clinical Practice Guidelines**

Another integral concept within the legal context is standard of care. Standard of care is defined as “the uniform standard of behaviour upon which the theory of negligence is based” (Gifsis, 2003, p. 490). Courts determine negligence based on what a reasonable person of ordinary prudence would do. Within the context of health law, standard of care can be discerned by evidence obtained by two different approaches: the use of expert witnesses and the use of clinical practice guidelines (CPGs).

Expert witnesses are frequently called upon to disseminate the current state of knowledge and general practices relevant to the issue in question. More specifically, they are retained to testify in relation to the following main issues: (a) the applicable standard of care, (b) causation (the link between the purported wrongful act or omission and the injury suffered by the client), (c) the assessment of damages, and, (d) a medical prognosis (Rosoff, 2001). The last three issues require the expert to apply their expertise directly to the question at hand; in other words, they are asked questions that will illicit responses based solely on their own belief and/or opinion. However, the assumption is that their belief and/or opinion will be based on some form of empirical observations. Asking an expert witness to speak on the applicable standard of care requires that they speak on behalf of their colleagues - how the majority of them would handle similar situations. The problem with this is that, in reality, experts are rarely able to attest to what constitutes the standard of care. Instead, their opinions typically amount to what they believe the standard practice of their colleagues should be, and in so doing, can discount other generally accepted procedures. Traditionally, the Courts compare the adequacy of care rendered against the standard practice in the relevant medical specialty; this has become known as the professional community standard. However, considering the expanse of medical practice, it is often difficult to subscribe to the school of thought that there could exist at least one community standard that the majority of practitioners have adopted.

Eddy (2001) proposed four conceptual steps to address these issues. The first step is recognition that health care takes place in a variety of settings and because of this, differing standards of care do exist. Thus, expert opinion should be proffered within this context. The
second step is realizing that a community standard most likely does not exist and that in fact the majority of physicians probably do not follow a single standard of care. The third step includes adopting the philosophy that truth is learned through empirical evidence as opposed to opinions of experts. However, when these opinions are proffered they should be based on the best available evidence. And the final fourth step is acknowledging that the subjective beliefs of experts should not be taken at face value; instead experts must be prepared to defend their position through an interpretation of the evidence.

In the case of medical malpractice, another factor to take into consideration when establishing a community standard of care is the state of scientific knowledge at the time of the incident as opposed to the time of the legal proceedings, “the conduct of physicians must be judged in the light of the knowledge that ought to have been reasonably possessed at the time of the alleged act of negligence” (Supreme Court of Canada, 1995). In this case (ter Neuzen v. Kom- a claim against a physician for infecting a patient with HIV through artificial insemination), the Court ruled that in 1985 it was not common knowledge that HIV could be spread in this fashion. The Court also noted that adherence to common practice would not automatically exonerate the physician from negligence if it could be determined that the common practice itself was grossly negligent.

The second approach to discerning the relevant standard of care is through the use of clinical practice guidelines (CPGs). CPGs are one of the outcomes of the evidence-based medicine (EBM) movement (to be discussed in more detail in a following section). CPGs are systematically developed statements outlining a recommended course of action a health practitioner should take relative to a specific clinical condition. However, the Courts need to be aware that not all CPGs are developed in a rigorous fashion. Many guidelines are still created through the traditional professional consensus approach (again, to be explained in a future section). Not only are clinical practice guidelines used to direct physicians and inform patients, they can also be utilized by health policy regulators and the Courts to assess appropriateness of care in a given situation. Often when guidelines are used in an attempt to prove the legal standard of care the need for expert witnesses is negated. However, if a guideline were being used to support the issues of causation, assessment of damages, or prognosis, it would not be considered strong enough evidence on its own without the supporting testimony of an expert witness (Rosoff, 2001).

Legal concerns over utilizing CPGs include whether the guideline is admissible, whether both the plaintiff and the defence can use guidelines to support their case, and eventually, how much weight (if any) would be afforded to the guideline. When a clinical practice guideline exists for the case in litigation, despite disclaimers made by its creators that the guidelines are merely “standard-setting”, if, and to what degree, a CPG becomes the legal standard of care is always at issue. However, due to the relatively early stage of clinical practice guideline
development, Courts are reticent to proclaim CPGs as the definitive answer to the legal standard of care. Instead, they may opt for the position that CPGs are evidence of customary practice, acceptable practice or as evidence of reasonable prudence (Rosoff, 2001). The assumptions surrounding standard of care include: (Eddy, 2001)

1. A standard of care exists and there is at least one way to achieve it.
2. That the majority of the professional community knows what the standard of care is and practices it.
3. There is no other practice that the majority of the professional community does that does not meet the standard of care.
4. The community standard of care should be consistent across communities (unless it can be explained by a lack of physical resources).
5. The community standard can be learned from experts.
6. Experts will present their beliefs without bias.
7. If experts disagree, the Courts have the ability to discern the “true” experts from the “false” experts (by evaluating their empirical evidence and their credentials).
8. There are some experts who know the true standard of care.
9. If there are a group of experts, they will be more accurate and less biased than individual experts.
10. All the practitioners who make up the community standard, or all the experts who provide testimony, all share the same philosophical perspective on what constitutes high-quality care.
Processes of Seeking, Understanding and Applying Evidence: Health Policy Context

Policy Process: The Policy Cycle

The policy process is the means by which policies come into effect. Here various social actors discuss and debate policy problems, priorities, and solutions. Although it is a collective, intellectual process, it is also a highly political one whereby power, influence, and interests interact with public needs. Over the years, attempts have been made to simplify the policy process by breaking it down into a number of discrete stages. For example, Pal (1992) outlined three elements or components to the policy process: (a) policy determinants, (b) policy content, and (c) policy impact, while Howlett & Ramesh (1995) expanded these concepts and logically linked each of their five subsequent stages to a corresponding phase of an applied problem-solving framework.

<table>
<thead>
<tr>
<th>Phases of Applied Problem Solving</th>
<th>Stages in Policy Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Recognition</td>
<td>Agenda-Setting</td>
</tr>
<tr>
<td>Proposal of Solution</td>
<td>Policy Formulation</td>
</tr>
<tr>
<td>Choice of Solution</td>
<td>Decision making</td>
</tr>
<tr>
<td>Putting Solution into Effect</td>
<td>Policy Implementation</td>
</tr>
<tr>
<td>Monitoring Results</td>
<td>Policy Evaluation</td>
</tr>
</tbody>
</table>

(Howlett & Ramesh, 1995)

While this policy cycle model appears to simplify the policy process, the inherent risk is that the policy cycle becomes interpreted as a systematic, linear progression towards policy development. This obviously does not correspond with reality where decision-makers are often faced with urgent problems that require immediate, ad hoc policy making strategies. Another criticism of the policy cycle model is that in practice, some cycles may be skipped, compressed, or even rearranged, looping back to previous stages when required. In addition, the model is criticized for lack of reference to causation – the "who" or "what" that drives policy from one stage to the next, along with its inability to highlight the complexities of the policy process. However, Howlett and Ramesh contended that, by expanding the scope of research questions to include inquiries into the actors, institutions, instruments and discourses within each of the policy stages, these limitations could be overcome. Their view was that policy research should be more than a general description of a policy context. Instead, it should strive to identify relevant variables and propose hypotheses about their relationships.

Agenda-setting

The first stage of the policy cycle is agenda setting. An agenda is a list of problems that governments focus on, while the actual methods used to recognize and prioritize these problems is the agenda-setting process. To begin with, problems must be defined. Problem
Definition is a process that involves recognition and definition of a particular issue of concern to the public. The problem must identify "a substantial discrepancy between what is and what should be" (Pal, 1997, p. 72). While some problems can be recognized objectively through "focusing events" such as crises, rises in indicators such as the unemployment rate, or through routine program monitoring and research, other problems are more subjective depending on the frame of reference assigned to it by the social or political actor(s). However, the mere existence of a problem does not precipitate policy; the problem must be recognized as an issue of importance to the general public for action to take place. Recognition can occur in numerous ways such as in response to a seminal publication, media reports, or the actions of social activists. A public problem is defined as one that, "either affects a substantial proportion of the public, offends or affronts some widely held public views or mores, or is itself the result of public policies" (Pal, 1992, p. 135). Since each policy actor will perceive the problem through a different lens, it must be carefully defined and articulated. Problem definition typically encompasses an in-depth examination of its characteristics such as causation (why the problem exists, e.g., poverty); severity (the seriousness of the problem, e.g., violence against women); the target population (those affected, e.g., children); and solutions (if available). Problem definition has a significant impact on the proposed solutions mainly through its causal explanations. Essentially, problem recognition and definition is, "the process of persuading others that a problem exists and that it should be addressed" (p. 77). It can be an arduous process marred by political posturing, metaphors and rhetoric. Labels placed on social problems can be strategically used to elicit strong moral responses (e.g., pro-life/pro-choice), to temper biases towards sexual orientation (e.g., homosexual vs. gay), or to soften reality (e.g., user fees vs. taxes). In addition, metaphors can essentially recreate an old problem into something perceived as new and sinister (e.g., "War on Terrorism"). The ultimate goal of problem definition is to position the problem high on the public agenda in order to draw out government response (Pal, 1997).

Problem definitions are then linked to specific policy goals. Goals are what the government is attempting to achieve, its aims, and its direction (Pal, 1992). Policy goals can be general in nature (e.g., to improve the health of all Canadians), or policy-specific (e.g., to decrease the rate of teenage pregnancies). General goals are endpoints in themselves, while policy-specific goals contribute to attaining the general goals. Often times policy goals precede problem identification. For example, by examining the Canada Health Act's principle of comprehensiveness, one problem that could be inferred is that not all medically necessary health care treatments are publicly funded.

The source of policy problems is referred to as policy determinants. These are the causal factors that are deemed responsible for generating a policy. These can include macro-level socio-economic factors such as the economy, culture, elections, public opinion, political parties,
or the media, and micro-level behavioural factors such as individual ideologies and special
interest groups. From this list it becomes obvious that policy-makers must respond to both the
objective conditions of the socio-economic environment in addition to the subjective discourse
that is common within society in general (Howlett & Ramesh, 1995; Pal, 1992). "The agenda of
politics or policy making . . . is an agenda which is established out of the history, traditions,
attitudes, and beliefs of a people encapsulated and codified in the terms of its political
discourse" (Howlett & Ramesh, 1995, p. 110).

Once a problem is identified, defined and its goals determined the problem must make
its way onto the policy agenda in order for it to receive attention. A policy agenda can have
two focuses: a systemic agenda and an institutional agenda. A systemic or public agenda is a
comprehensive list of all the public issues identified by the government that are within its legal
jurisdiction to act upon. This becomes the agenda for public discussion. Alternatively, an
institutional or formal agenda lists the issues selected from the systemic agenda that are
targeted for action by the government. In order to end up on the institutional agenda, problems
have to go through a succession of four stages. First, the public issue is initiated; next, the
solutions are specified; and then support for the issue is expanded. Finally, if the issue is
successful, it enters into the institutional agenda (Cobb, Ross & Ross, 1976, as cited in Howlett &
Ramesh, 1995, p. 112). It is worthy to note that either private interest groups or members of the
government can initiate this public policy making process. In fact, the catch phrase "iron
triangle" was coined for the coalition of politicians, special interest groups, and members of the
media that join forces to champion their cause onto the institutional agenda (Rist, 1994).

Agenda setting involves the processes of discussion, debate, and persuasion amongst
policy actors. Which actors become involved in the policy process will depend upon the context
of the problem. Technical issues are best handled by substantive experts, while problems with
any ethical, social or political implications generally require a larger, heterogeneous forum of
participants. Agenda-setting also involves creating an image for the problem as it is well
recognized that how a problem is framed not only influences those who will support a policy
position but it also impacts the policy activities that follow (Howlett & Ramesh, 1995).

Agenda setting also involves prioritizing health service options. One local example can be found
in Mitton, MacKenzie, Cranston and Teng (2006). These British Columbian health care
administrators recognized that the historical allocation method of basing a program's current
budget on the previous year's allocation (with some adjustments contingent on the current
financial and political climate) was not the most effective method of ensuring value for dollar
spent. Their survey of key decision-makers confirmed that prioritization is often an ad hoc process
that, in terms of resource allocation, typically results in "the squeaky wheel getting the grease".
With the goal of being "more transparent and defensible", these administrators developed a
formalized priority-setting model. This process consisted of developing a set of decision making
criteria based on organizational values and criteria uncovered in the literature. These criteria were validated and weighted and then applied against each program option. A notation was also made as to whether the criterion assessment was based on expert opinion or evidence. A point of interest though was that equal weight was assigned to both.

**Policy Formulation**

The second stage of the policy cycle is policy formulation. Here options are reviewed for solving the identified problem. Sometimes the solutions are proposed in conjunction with the problem identification and definition phase; while at other times, options are conceptualized later in the policy cycle during policy formulation. Either way, rarely are options comprehensively assessed, taking into consideration their safety, effectiveness, costs, and ethical, legal, and social implications (Kazanjian & Pagliccia, 1998). In addition, the public's input is often restricted. Instead, state actors investigate the range of policy options and only those deemed feasible are entertained. Therefore, the key functions of this stage are, “defining, considering, and accepting or rejecting [policy] options” (Howlett & Ramesh, 1995, p. 122). However, the literature stressed that searching for policy options is by no means an orderly process. Instead, policy-makers are faced with multiple constraints delineated as either substantive or procedural. Substantive constraints are specific to the problem itself and tend to focus on the larger issues. With these problems, solutions are not simple and may in fact be elusive. Or, if alternatives are identified, they often carry such substantial consequences that they are not considered as viable options. Procedural constraints, on the other hand, are limitations imposed by the actual procedure(s) involved in implementing a policy option.

**Decision Making**

The third stage of the policy cycle is decision making. The process of policy decision making is one of the most researched areas of policy science. Early decision making models were borrowed from the fields of business organization and public administration and range from systematic and comprehensive to ambiguous and unpredictable. Public policy decision making, however, incorporates several essential elements. First, it is recognized that decision making is not an entity separate from the other stages of the policy cycle. Instead it builds upon previous stages and requires the application of logic in deciding upon various policy options. Second, policy decision making is a highly political process, not a technical exercise. Third, as the policy process continues, the number of policy actors decreases to the point where most non-state actors are excluded (including those from other levels of government). The policy actors that remain are only those that have the authority to impose the policy. Fourth, it is recognized that individual policy actors are constrained in their actions by a multitude of rules and regulations that limit their freedom. Instead, “action channels” – “a regularized set of procedures for producing certain types of decisions” (Howlett & Ramesh, 1995, p. 138) are established to process the routine policies. However, decision-makers are still afforded the
freedom to determine “the best” policy option depending upon the context of the specific situation.

By the 1960s, two public policy decision making models were predominant: the rational and the incremental models. The rational model, which has its ideological roots in early Greek and European civilizations, with its more recent application in business organizational theory, conceptualizes decision making occurring as a prescriptive, sequential process. First of all, a problem is identified and defined, its goal created and policy alternatives identified. Then, consequences of each policy alternative are predicted and their probabilities of occurrence are estimated through the use of modeling. The consequences include costs and benefits along with positive and negative impacts. Next, the policy alternatives are ranked in order of preference and the strategy that best addresses the problem, or is the most cost-effective, is chosen and implemented. And finally, the policy is evaluated to determine the need for change or revision. Rationalist theories subscribe to the belief that scientific and rational methods should be used to solve policy problems and that this process should incorporate in depth information gathering on both the problem itself and the policy options. Of note, however, is the fact that rational decisions are not synonymous with reasonable decisions. A rational decision is descriptive of the process utilized to come to a decision, whereas a reasonable decision is judged as such by its appropriateness as a solution to a problem. Rationalism is essentially a Western concept, with other cultures relying on other systems of knowledge such as wisdom, intuition, or signs. In addition, rationalism presumes that there is a linear, systematic, self-conscious, purposive and efficient pattern to thoughts (Pal, 1997).

The rational decision making model was later expanded into the rational–comprehensive model whereby policy analysts argued that the only way the rational approach would be effective would be by investigating all policy options and their relevant costs. However, the criterion of comprehensiveness was difficult to achieve due to the realities of political and institutional constraints within the policy world. Other criticisms included the cognitive limits of decision-makers in regard to being able to identify all policy options, the implausibility of anticipating consequences of each option in advance, and the difficulty in comparing options to each other, due to unique nuances of each option under specific contexts (Howlett & Ramesh, 1995; Pal, 1997, Pal, 1992).

When support began to wane for the rational model, policy analysts sought a model that more accurately matched the actual processes of government decision making. The incremental decision making model (developed by Yale University political scientist Charles Lindblom) is a political model applied to public policy; it is characterized by bargaining and compromise amongst self-interested policy-makers. Lindblom’s view was that decision-makers developed policy through an analysis of previous decisions, taking incremental steps away from the status quo. The key features of the incremental model are: (Howlett & Ramesh, 1995)
1. Only a few key policy alternatives are considered.
2. The policy alternatives must be familiar and should not differ substantially from the status quo.
3. Instead of first identifying relevant values, and then designing means to promote them, incremental policy decision making promotes analysis of policy goals and other values alongside empirical evidence of the problem.
4. A greater emphasis is placed on solving the problem as opposed to achieving a positive goal.
5. Policy options are discovered through trial and error.
6. Only some of the important potential consequences are analyzed.
7. The analytical work is divided amongst many partisan policy actors.

Lindblom also did not agree with the rational model's separation between means and the end arguing that in some policy areas ends were inseparable from means and that the goals to be achieved would depend upon the means that were available to the policy-maker. In effect, the incremental model views decision making as a practical problem-solving tool rather than a political process to attain impressive goals.

However, incrementalism did not escape the critics' wrath. Criticisms included the model's lack of attention to goal orientation and its overly conservative response to the prospect of dramatic policy changes. In addition, the incremental model has been called undemocratic in that policy is made by only a few select senior policy-makers. And finally, its vision has been criticized as being somewhat short-sighted because it does not advocate systematic searches for new policy alternatives. This perspective may result in adverse consequences for society (Howlett & Ramesh, 1995; Pal, 1997, Pal, 1992).

In the early 1980s the debates on the supremacy of rationalism vs. incrementalism slowly gave way to the realization that there were multiple decision making styles and that the context of the policy decision determined which specific style was adopted. Simon and March (March, 1982) studied the behaviours of skilful decision-makers and recognized that decision making was not a logical, orderly process due to the cognitive limits of the decision-maker in addition to the numerous constraints they face, thus modifying the theory which became known as bounded rationality.

John Forester (1984) built upon Simon and March's bounded rationality theory by identifying five conditions that were common to all decision making models: (a) agent, (b) setting, (c) problem, (d) information, and (e) time. From there he predicted that the degree of complexity within each condition would dictate the practical strategies used to solve a problem. For example, at the simplest level (the rational comprehensive approach as described above), there would first have to be a limited number of agents (decision-makers) involved in the policy process. Second, the organizational setting would have to be simple and separate.
from the influences of other policy actors. Third, the problem would be well defined so that its scope, values and consequences were fully understood. Fourth, information on the problem and policy solutions would be complete, accessible and comprehensive. And, finally, time would not be a factor in that there could not be time constraints placed on the policy decision making process so that decision-makers would investigate each policy alternative. With this scenario, "rational problem solving or optimization through available algorithms and solution techniques" (p. 25) would be the best practical strategy.

However, decision making contexts are not typically this simple. Forester’s model took the five conditions and placed each of them along a continuum from the simplest context (as described above) to the most complex. For example, the condition agent goes from a single decision-maker all the way up to groupings of multiple planners, and administrators. Similarly for the condition of information, this can range from perfect information up to misinformation or ideological information. The result is a taxonomy of five specific styles of decision making strategies, (a) optimization/solve, (b) satisfice [sic]/hedge, (c) network/search, (d) satisfice bargain/increment, and (e) organize/democratize. Each type of decision making strategy would also be based on five types of "boundness": (a) comprehensive (unbounded), (b) cognitive limits, (c) socially differentiated, (d) pluralist (Lindblom), and (e) structurally distorted/political-economic (Forester, 1984).

Howlett and Ramesh (1995) simplified Forester’s model by grouping the variables of agents and setting together into the category of policy subsystem, and placing the conditions of problem, information, and time under the concept of constraints. Subsequently, their model depicts four styles of decision making: (a) incremental adjustment, (b) satisfying search, (c) optimizing adjustment, and (d) rational search.

Table 2.2 Decision Making Strategies

<table>
<thead>
<tr>
<th>Complexity of the Policy Subsystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity of Constraints</td>
</tr>
<tr>
<td>High</td>
</tr>
<tr>
<td>Low</td>
</tr>
</tbody>
</table>

(Howlett & Ramesh, 1995, p. 148)

The decision making style of incremental adjustment, which assumes a highly complex policy subsystem, coupled with a high level of constraint, is comparable to Lindblom’s incremental decision making model (and Forester’s pluralist model). Alternatively, when both the complexity of the subsystem and the level of constraints are low, a rational search is feasible for new and different policies. Satisfying search is an option when there is a low complexity to the subsystem paired with a high level of constraint. In this context, expectations are lowered from finding the
optimal solution to a satisfactory one. While the fourth style - optimizing adjustment - which operates within a complex policy subsystem with low constraints, is conducive to achieving the optimal policy adjustments.

**Policy Implementation**

The fourth stage of the policy cycle - policy implementation - is the act of putting policy decisions into practice. However, despite the most well thought out implementation objectives, this stage is invariably hampered by a host of contextual factors. These can include restrictions due to the technical complexity of the program, its cost, its interdependency with other programs, the characteristics of its target audience, the diversity of problems targeted by one program, the extent of behavioural change expected from the policy, changes in social conditions, availability of the new technology, changes in political parties, or changes at the bureaucratic administrative level (Howlett & Ramesh, 1995).

Public policies are implemented through the application of policy instruments to social problems. Policy instruments (also known as policy tools or governing instruments) are the actual methods used to solve the identified issues. In practice, there can be consensus between policy actors in regard to the specific problem and its goals as these issues may be well understood and accepted in the policy field. Instead, it is often the “how to” or, the policy instrument, which is frequently contested. In fact, advocacy groups may initiate the policy process with a specific policy solution already in mind, then proceed with trying to convince the government that a problem exists, that it has the specific characteristics that they have defined, and that their solution would be the best policy response.

A policy instrument impacts private citizens and carries with it the assumption of the government's power and authority in doing so. Inherent in this process are societal values, specifically the concept of legitimacy - an ethical concept depicting “right” and “wrong” (or, “fair” and “unfair”). The perceived level of coercion inherent in a policy instrument can influence the legitimacy of the policy.

There are four classifications of policy instruments at a government's disposal, however, similar to problems and goals, public policies may contain multiple policy instruments. The first are nodal instruments (viewed as the least coercive), which involve passively providing, and collecting information without imposing rewards or threats. This voluntary process requested of the public is based on the assumption that, "motivations are strong enough that individuals can be relied upon to achieve policy goals once apprised of new information" (Pal, 1992, p. 145). Here, the classic example is assuming that the public will be motivated to prevent the spread of HIV through the practice of safe sex. The second classification is treasure, where money (an expenditure instrument) is used to implement public policy. This can be done through the use of numerous options under the categories of grants, subsidies, and purchases. The third classification is authority, or the state's ability to prohibit and permit actions. Instruments that rely
on authority are the most coercive. These instruments include public regulation (e.g., licensing, criminal laws, and rent controls), self-regulation (e.g., professional associations), and sanctions (e.g., punishments such as suspensions, levies, and physical punishment). And, the fourth policy instrument is organization or the use of the state’s own resources such as land, buildings, equipment and personnel to implement policies. While the three previous instrument categories – nodality, treasure, and authority – require governments to implement policy through intermediaries, organizational instruments permit governments to act directly on policy changes (Pal, 1992).

How policy instruments are chosen is often based on the subjective preferences of the decision-makers relative to their professional background, institutional affiliation and cognitive make-up. The instrument choice depends upon the specific features of the instrument, the characteristics of the policy problem, government’s past experience with both the instrument and the problem, and society’s anticipated reaction to the instrument (Howlett & Ramesh, 1995).

**Policy Evaluation**

The final stage of the policy cycle is policy evaluation. Conceptually, policy evaluation is the process of determining how the policy works in practice, the methods used, and the objectives met; it is grounded in the rationalist perspective. Once a policy has come into effect, the requirement or need to evaluate it can present itself from different perspectives. First of all, governments may elect to evaluate the delivery of services provided within the policy program to determine the “value for money” for goal achievement. This is termed administrative evaluation and requires complex procedures of data collection, cost comparisons, and outcome measurements.

Next, policy evaluations can originate outside of the government by social actors such as special interest groups, the media, professional associations, or anyone with an interest in a specific policy. These evaluations are typically not systematic or methodologically sound; instead, they lean towards their ideological biases. This form of evaluation is generically termed political evaluation and carries with it the objective of either supporting or challenging a policy. Depending upon the weight of the evaluation in regard to its social influence, political evaluations can force the government back to previous stages of the policy cycle to review and possibly revise their policy. Other forms of political evaluations include elections and policy consultations. In regard to elections, this form of political evaluation is typically representative of the voter’s opinion of the government’s entire policy platform, as opposed to one specific policy. Policy consultations on the other hand are methods used by the government to seek input into policy. These can take the form of meetings, task forces, forums, and public hearings. The goal is to ascertain the view of various policy subsystems and members of the public in regard to the specific issue under question (Howlett & Ramesh, 1995).
And finally, the third form of policy evaluation is judicial evaluation. As previously described, the purpose of a judicial review is to assess the manner in which a government implemented a program with the focus on identifying possible incongruencies between government actions and constitutional provisions or established administrative standards of conduct and individual rights. Canadian Courts concentrate on whether a government agency acted within its powers or jurisdiction, taking into consideration the principles of natural justice, without acting in an impulsive or arbitrary fashion. The focal point is on issues or errors in law as opposed to errors in fact (Howlett & Ramesh, 1995).

Evidence-Based Decision Making: Health Technology Assessment

One method that policy-makers call upon to aid their decision making and evaluation processes is health technology assessments (HTA). HTA is a multidisciplinary form of evaluative research that evolved from the broader era of technology assessment of the 1950s and 60s. During this period technology assessment was primarily utilized as a decision making tool for purchasing products and developing processes. Engineers, technicians, and economists employed systematic methodologies in areas such as offshore oil drilling, pesticides, automobile pollution, nuclear power plants, supersonic airplanes and artificial hearts. These methods included systems analysis, cost-benefit analysis, consensus methods, engineering feasibility studies, clinical trials, market research, and technological forecasting. At this time the public unquestionably accepted the latest technological advancements. By the 1980s there was a rapid growth of technology assessment as the purpose of evaluation shifted towards an examination of its consequences. In light of the thalidomide disaster, there was a movement toward tighter control of technology. Within institutionalized and government settings health technology assessment became a separate evaluation entity. This decision framework, which incorporated the attributes of equity and utility and was grounded in the principles of social justice, provided a defensible argument to even the most difficult policy decisions.

The concept of health technology is broadly defined and includes drugs, devices, medical and surgical procedures, and organizational and administrative support systems. The field of HTA not only examines the safety, efficacy and effectiveness of technology, but also is a process for policy research in that it examines both the short-term and long-term consequences of technologies (Kazanjian, Bassett & Savoie, 2001). HTA is grounded in concepts and methods from epidemiology, health services research, bio-statistics, health economics and ethics. The variety of qualitative and quantitative methods utilized in HTAs produce scientifically based effectiveness evidence that can be utilized by health care administrators to guide technology decisions. The evidence supports the development of standards, policies and guidelines such as clinical practice guidelines and evidence-based medicine. HTA can be applied to the levels of prevention, screening, diagnosis, treatment, rehabilitation and palliation. In addition, depending upon policy requirements, HTA can be employed throughout the entire
technological lifecycle (stages of diffusion) from the conceptual stage, experimental stage (pre-clinical), investigational stage (clinical), established stage, through to the outmoded or obsolete stage (Kazanjian & Pagliccia, 1998). HTA provides decision-makers with an approach to address the population perspective goal of equity and utility through an assessment of the safety, effectiveness, financial, ethical, legal and social implications of a proposed technology. Results can be used to influence the development and diffusion of technology through governmental actions such as funding of specific research projects, regulation of technology marketing and facility licensing, and the allocation of public funds for technology adoption. However, in reality, most technologies are accepted into the health care forum before efficacy or effectiveness evaluations can be completed.

Health Technology Assessment Framework

Strategic HTA is a broadly based discipline that evaluates technology within the contexts of community dynamics, institutional interests, and hard science. In addition to describing the technological impact, this framework also synthesizes the social dynamics inherent to the proposed technology by critically examining issues of politics, professional authority, community beliefs and values, power and dominance. It provides a venue to fully understand the health problem and the consequences of the proposed technological solution (Kazanjian & Pagliccia, 1998).

While many different HTA frameworks exist, (Alberta Heritage Foundation, 2000; National Information Centre on Health Services Research & Health Care Technology, 1998), the framework selected for this research was developed by the British Columbia Office of Health Technology Assessment (BCOHTA) (Kazanjian, 2004; Kazanjian, 2003; Kazanjian, et al., 2001; Kazanjian & Pagliccia, 1998). This comprehensive (or, “expanded”) HTA framework goes beyond the efficacy/effectiveness and economic perspectives to incorporate the epidemiological and social contexts of the technology under question (Table 2.3).
### Table 2.3 Comprehensive Health Technology Assessment Framework

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Indicators (examples)</th>
<th>Target/Goal (examples)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Epidemiological</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Population at risk</td>
<td>Health services utilization/access</td>
<td>Reduced health deficits</td>
</tr>
<tr>
<td>- Population impact</td>
<td>Mortality, morbidity</td>
<td>Reduced burden of illness</td>
</tr>
<tr>
<td><strong>Economic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Economic issues</td>
<td>Cost-effectiveness analysis</td>
<td>Recognition of allocative efficiency.</td>
</tr>
<tr>
<td>- Cost-benefit analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Social</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Ethical, legal and political</td>
<td>Social context</td>
<td>Balanced gender participation in decision making.</td>
</tr>
<tr>
<td></td>
<td>Legal framework</td>
<td>Development of legal perspective.</td>
</tr>
<tr>
<td><strong>Effectiveness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Technology assessment</td>
<td>Comprehensiveness of scientific evidence.</td>
<td>Increased understanding of conflicting interests.</td>
</tr>
<tr>
<td>- Role of scientific evidence</td>
<td>Source of scientific evidence</td>
<td>Improved relevance of evaluative research.</td>
</tr>
<tr>
<td>- Quality of scientific evidence</td>
<td>Convergence of scientific evidence.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first three dimensions of BCOHTA’s framework—epidemiological, economic, and social—are the descriptive elements of the health problem framed within the social and environmental contexts. The fourth dimension—effectiveness—provides information on the strength and quality of the empirical evidence through systematic review and critical appraisal methods. Each dimension has clearly defined, measurable indices that are synthesized with other qualitative measures to present composite measures that will describe the specific policy issues of importance. These social indicators are a way of measuring social conditions and changes over time within a specific population. BCOHTA’s framework facilitates a broader problem-solving approach in order to ensure a thorough understanding of many facets of the health problem at hand. In addition, the framework can also incorporate an examination of the consequences of alternative courses of action.

**Epidemiological context.** Epidemiology is the study of the distribution and determinants of diseases and injuries in human populations in relation to the variables of person, place, and time. The variable of person examines specific disease characteristics related to sex, gender, age, race, marital status, and socio-economic status; place is the geographical distribution of the illness, while time examines the variance in time of the disease occurrence. The epidemiological context of an HTA examines the population at risk and the population impact. The goals are first to determine the recipients of the technology and then to uncover both qualitative and quantitative empirical evidence to describe this population in epidemiological terms. Population at risk looks at the magnitude of the health problem and defines it based on
empirical evidence on the epidemiological measures of incidence and prevalence. Population impact seeks to understand the burden of illness and examines the expected and unexpected consequences of a health technology on the population at risk.

**Economic context.** Health care technologies can be evaluated in regard to the economic impact they will have on the health care system. HTA methodologies can accommodate inquiries into microeconomic issues such as the cost of the proposed technology up to macroeconomic concerns such as the impact of new technologies on national health care costs, the effects of reallocation of funds between programs, to the effects on regulatory policies and health care reform (National Information Centre on Health Services Research & Health Care Technology [NICHSR], 1998). The economic context of the HTA framework examines inputs into a health care program such as its direct costs to the health care system, its indirect costs in regard to morbidity and mortality, and its intangible costs such as pain and suffering, then compares them to the health outcome of interest. These outcome measures can be summarized as clinical outcomes by way of a cost-effectiveness analysis (e.g., number of life-years saved); as a monetary value of different health effects through a cost-benefit analysis; or, in measures that also incorporate the quality of life through cost-utility analyses (e.g., quality adjusted life-years – QALYs) (Kazanjian, 2003).

**Social context.** The social context of the HTA framework examines the ethical, legal and political concerns of a technology under question. More specifically, social impact analysis contributes to the understanding, explanation, and prediction of potential effects of technology on social systems such as individuals, population groups, health professionals, communities, organizations and institutions. It is important to examine these social actors as the diffusion of a new technology can have implications far beyond the obvious ones. Social indicators are qualitative measures that can be assessed at the level of the individual/family, unit, community, organization, or system level. However, these social systems are not clear-cut dimensions that can be examined in isolation. In reality, these social units interact and impact the effects of the technologies (Kazanjian & Pagliccia, 1998).

Emphasis is placed on evaluating the anticipated effects of new technologies within the contexts of medical ethics and social justice. From an ethical standpoint, this would be assessing the impact on the tenets of autonomy, beneficence, non-malfeasance, and justice. Autonomy refers to the ability of the patient (or their family) to retain some control over their health care; beneficence is a true benefit to the patient from the proposed technology; non-malfeasance is the potential for a technology to do harm; while justice relates directly to the philosophy of universal public health coverage balanced against expensive new technologies and shrinking health care budgets (Kazanjian & Pagliccia, 1998). Equity is another important social indicator to examine to ensure that all those that would benefit from the technology would have access to it. And finally, the political implications of the technological implementation should be
reviewed. Exactly which dimensions of the social context are examined, and the weight placed on each dimension, would depend upon the health issue of concern. However, in the end, the goal of examining the social context within an HTA framework is, "understanding factors within the socio-political context which underlie the demand for a technology" (Kazanjian, et al., 2001).

**Effectiveness context.** The final dimension of the comprehensive HTA framework examines the effectiveness of the technology through an evaluation of its performance, clinical management, safety, and health outcomes. Effectiveness differs from efficacy in that, while they both convey how well a technology functions, “efficacy refers to the benefit of using a technology for a particular problem under ideal conditions [e.g., within a randomized controlled trial]... [while] effectiveness refers to the benefit of using a technology for a particular problem under general or routine conditions [e.g., a family physician’s office]" (NICHSR, 2000, p. 6).

An assessment of a technology’s effectiveness is accomplished through an examination of the available evidence utilizing the objective scientific methods of systematic review and critical appraisal. Technical properties that can be evaluated include performance characteristics, while safety considerations involve a judgment on the acceptability of risk (the probability of an adverse event and its severity) balanced against the proposed benefits when used within a given situation (NICHSR, 2000). Rules of evidence are an integral part of a HTA framework where strict criteria are applied to appraise the quality of a study, and to draw conclusions on the strength of evidence of a body of knowledge that will be used to facilitate decision making.
Processes of Seeking, Understanding and Applying Evidence: Health Care Context

Peer Review Process

Peer review is, "a resource-intensive process relying on considerable, chiefly volunteer, effort to evaluate manuscripts for publication and craft objective and constructive reviews" (Weber, et al., 2002, p. 2790). Medicine has held firm to the belief that peer review, "raise[s] the quality of the end product and provide[s] a mechanism for rational, fair and objective decision making" (Jefferson, et al., 2002a, p. 2784). The purpose of peer review is to filter out incorrect or inaccurate information and to improve the accuracy and clarity of published reports (Jefferson, et al., 2002a). Although peer review dates back to the 18th century (Kronick, 1990), until more recently, it has essentially gone uncontested as being the premier approach used to critique and disseminate medical evidence. However, this perspective began to change in 1985, first with the publication of Stephen Lock's book on peer review entitled *A Difficult Balance*, and second, with the creation of the *International Congress on Peer Review in Biomedical Publications* by the *Journal of the American Medical Association* (JAMA) (Rennie, 2002). Since its inception, JAMA has held five conferences that have endorsed the goal of supplanting opinion with research on editorial peer review. In essence, these two events sparked a whole new field of scientific inquiry.

Research into peer review has begun to examine the issue of open review vs. blinded or masked review. One randomized study (van Rooyen, et al., 1998) found that blinding the reviewers to the author, or revealing the identity of the reviewer, made no difference to the quality of the review, the time it took to complete the review, or the eventual publication decision. However, these results contradicted a previous randomized trial that found the quality of the review improved with the blinding of the reviewers (McNutt, et al., 1990); and with a more recent study that suggested that open peer review biased reviewers towards authors from prestigious institutions (Ross, et al., 2005). A systematic review of seven studies (that included van Rooyen & McNutt) concluded that the research to date has produced inconclusive evidence on the effect of blinding/masking (Jefferson, et al., 2002a). The review also showed that, (a) the hypothesis that peer reviewer training would improve the quality of the reviews was not supported, (b) there was no proof that using a statistical checklist improved the reviewers' submissions, and (c) there was no evidence of the effects of peer review on the study's validity (however, the authors did acknowledge that bias may have impacted these results). There were a couple of positive outcomes from the systematic review. First, the researchers determined that there might be a link between peer review and study report quality; and second, that there was no evidence of bias by reviewers towards unconventional therapies (although, again there may be a bias introduced by a response rate of 61%). However, the authors concluded that, due to methodological weaknesses, the internal and external validity was difficult to assess.
Weber, et al (2002) examined authors' satisfaction with the peer review process and, not surprisingly, found that satisfaction was positively correlated with manuscript acceptance and inversely correlated with manuscript refusal; in both cases, the quality of the review was not associated with author satisfaction. Another investigation, into whether editorial feedback provided to reviewers improved the quality of subsequent reviews, failed to show a difference (Callaham, Knopp & Gallagher, 2002), while a study exploring the characteristics that identify a "good" reviewer, along with the length of time required to produce a "good" review, failed to identify characteristics of high-quality reviewers. However, the researchers did determine that spending more than three hours on a review did not increase its quality (Black, et al., 2002). And finally, the results of a study examining the inter-reviewer reliability suggested there was a relatively low level of agreement among peer reviewers, with the lowest consistency among the expert reviewers and the highest between editors and editorial board members (Scott, Martin & Burmeister, 2005).

It appears that very little is known about the effects of the peer review process. Research conducted since 1985 showed a wide range of study questions and outcomes, which suggests that the expectations for peer review are diverse, that its true effects have not been determined, or the aims of peer review have not been properly identified (Jefferson, Wager & Davidoff, 2002b). Researchers have proffered suggestions on how to bridge this gap. For example, it has been suggested that journal editors should assume a greater role in mitigating the effects of manuscript rejection by encouraging authors to use criticism to improve their research and writing (Weber, et al., 2002). Jefferson et al. (2002a) suggested that more questions could be answered if the focus turned to examining rejected manuscripts as opposed to the accepted ones. Horton (2002) advocated that journal editors develop guidelines that published authors must follow in regard to addressing criticisms received through subsequent letters to the editor. While Jefferson et al. (2002b) proposed a framework for identifying, defining and assessing quality indicators.

It is somewhat ironic that peer review, (acknowledged as an extension of the scientific process), has been blindly accepted as the gold standard for critiquing the methodological rigor and validity of studies, yet it does not hold its own reviewers (and processes) accountable to standards similar to those imposed on the authors. Even Dr. Drummond Rennie, the Deputy Editor of the JAMA, in response to the research submissions for the 2002 Congress, remarked: "once again, . . . we publish studies that fail to show any dramatic effect let alone improvement, brought about by editorial peer review" (Rennie, 2002, p. 2759); and admitted that, if this process were to be introduced today, it would be a hard sell based on the lack of evidence. However, despite this, journal editors' conviction towards the utility and importance of peer review remains unwavering. Rennie proffered the suggestion that perhaps the wrong tools were being used to study the wrong factors.
Consensus Development

Consensus development is another form of peer review in which a decision making process is used to reach agreement on a technology’s efficacy, safety, and conditions for use (IOM, 1990). Historically, medical practice has been guided by tradition, experience and consensus-based guidelines. Consensus methods are qualitative approaches that are applied to scientific evidence for the purpose of reaching agreement. The process is implemented when either: (a) there is a lack of scientific evidence, or (b) when the available evidence is contradictory. The aim of consensus methods is to determine the level of agreement amongst experts and lay people pertaining to the issue under investigation.

The National Institutes of Health (NIH) Consensus Development Conference (CDC) is held out to be the gold standard of consensus methods. The CDC was designed around three models: (a) the judicial process, (b) the scientific meeting, and (c) the town meeting (Andreasen, 1988). The conference simulates a judicial process in that the “jury” (the conference panellists) hands down a “verdict” (a consensus statement) after weighing the “evidence” (the scientific findings) presented by “witnesses” (the expert speakers) (Wortman, Vinoku & Sechrest, 1988). Incorporating the concept of the scientific meeting, where experts discuss their work with peers, and by adding the town meeting process, where the forum is made public and the audience is invited to participate, assists in augmenting the underlying judicial process.

The consensus development process begins with the selection of the conference subject. Within the United States, the NIH or other public health service agencies typically suggest a topic based on the following criteria: (a) public health importance, (b) controversy or gap in knowledge, (c) available scientific information to answer the conference questions, (d) amenable to clarification on technical grounds, (e) health care cost impact, (f) preventive impact, and (g) public interest (IOM, 1990). After the topic is decided, the next step is the selection of participants; these may include clinical experts, scientific experts, and non-experts. Again, typically these experts are recommended by the NIH planning committee based on personal acquaintance and professional reputation. Ideally, the criteria for the panellists is that they be neutral – those who have no publicly stated position on the conference topic – and have experience in evaluating evidence and working within a collaborative process (Wortman, et al, 1988). The speakers are also usually recommended by NIH and are asked to present evidence on the technology’s safety, efficacy, effectiveness, and service requirements. The panellists are then presented with a list of key questions that focus on the efficacy, risks, clinical applications, and directions for future research. These are addressed throughout the conference based on evidence presented by the key speakers.

The outcome of the CDC is the Consensus Statement. This is a compilation of each of the sub committee’s conclusions and recommendations based on their respective consensus questions. Once consensus is reached, this statement is disseminated to the respective parties.
such as the media, the *Journal of the American Medical Association*, federal agencies, health care organizations, and various alternative journals.

Since its inception, the NIH Consensus Development Conference has been criticized for its lack of a systematic process and criteria for selecting topics, and the absence of a transparent search strategy for scientific evidence (Lomas, 1991; Wortman et al., 1988). While computer-based data sources are utilized, the CDC does not use pre-defined criteria to establish what constitutes relevant research. These omissions often leave the consensus process vulnerable to political power and personal bias and render them un-replicable and indefensible. In addition, topics are often selected that have insufficient evidence to support consensus.

NIH has also been accused of selection bias in regard to the types of questions asked. Questions pertaining to controversial aspects of technologies, such as social, ethical, legal, and economic issues, are often overlooked. Instead, it is felt that the conference should be limited to biomedical science and the pursuit of a "technical consensus" (Wortman et al., 1988). This perspective is congruent with the NIH guidelines that indicate, "the primary focus of a consensus development conference is on technology's clinical applications" (IOM, 1990, p. 42). Again, questions can also be omitted if there is not enough scientific evidence to support the consensus process. Other criticisms included an excessive agenda with too many questions being asked in insufficient time.

Selection bias can also be an issue in the participant selection process. Past conferences have been criticized for "stacking the panel" in favour of the clinical status quo or NIH-supported recommendations (Wortman et al., 1988). Participants have historically been assembled from narrowly focused segments of the biomedical community and are often personal or professional acquaintances of NIH staff. Few, if any, criteria exist for identifying potential conference participants.

As previously mentioned, NIH determines the scope and nature of the literature search. Criticisms have been made that, (a) the data provided to the panellists has been incomplete, (b) the selected studies have contained methodological flaws, (c) equal weight is awarded to each research design, and (d) consensus findings are often generalized to inappropriate populations (Wortman, et al., 1988).

In regard to the development of consensus, critiques included, (a) not enough time allocated to developing consensus, (b) use of various decision analysis models confusing the issue, (c) the CDC process not facilitating alternate viewpoints, and (d) the written consensus statement not containing references to the scientific literature reviewed. There have also been issues in regard to the quality of the scientific writing, although previous attempts to incorporate scientific writers into the panel were met with disapproval from the participants (Wortman et al, 1988).
Some other general criticisms included the assertion that NIH conferences were often ill
timed and perhaps even irrelevant (Riesenberg, 1987), that they were a social process as much
as they were a scientific process, and that this inherent social nature could lead to imperfect
scientific recommendations (Sniderman, 1999). In addition, when research evidence was
scarce, subjective judgments and interpretations were often substituted. In the past, consensus
conference planning procedures had been described as "haphazard and inadequately
structured" (Wortman et al., 1988, p. 49).

Implicit in most dissemination plans is the assumption that distribution of information will
result in a change in practice (Kanouse & Jacoby, 1988). However, research into the impact of
consensus statements shows otherwise (Kosecoff, et al., 1987; Lomas, et al., 1988; Lomas, 1991):
"Providing health care practitioners with clinically relevant information will not necessarily
change their behaviour" (Kanouse & Jacoby, 1988, p. 29). To influence practice, there must be a
better understanding of, "other factors that determine whether change will occur, such as
practitioners' motivation to change, the context in which clinical decisions are made, and how
information is presented" (p. 27). In order to effect change, the process must deliver a
scientifically sound, timely, and clinically relevant consensus statement to the appropriate health
professionals who are willing, and able, to adopt it (Kosecoff, et al., 1987). In Kosecoff's study,
the researchers found that, despite the apparently sound recommendations contained within
the consensus statements, coupled with NIH's moderate attempts at disseminating the results,
medical practitioners still did not adopt the conference's recommendations. Often, however,
change in practice is dependent upon factors beyond a practitioners' control, such as access
to the required technology.

**Evidence-Based Medicine**

Historically, physicians freely introduced diagnostic and treatment strategies into their
daily practice without much regard for concepts such as observation, comparison of outcomes
or potential harms and risks. In fact, it was not until the last couple of decades that the medical
profession slowly began to develop an underlying value system cemented in the principles of
scientific inquiry. This philosophical change from "the healing arts" towards an exact science
was precipitated by a number of factors. First of all, computers greatly facilitated the advent of
outcomes research by their sheer computing capacity; also, significant medical mishaps
(e.g., the thalidomide tragedy), also prompted the search for more stringent rules and
regulations (Rosoff, 2001).

Evidence-based medicine (EBM) began as an approach to handle and process a
sudden overload of medical information. David Sackett, the founding father of EBM, defined it
as, "the conscientious, explicit and judicious use of current best evidence in making decisions
about the care of individual patients" (Sackett et al., 1997, p. 71). EBM is a by-product of the
clinical epidemiology movement introduced by Gordon Guyatt in the late 1970's (Sackett,
2002). Theoretically, evidence-based-medicine usurps historical approaches such as authority, tradition, physician's personal experience, and consensus-based judgments by substituting systematically analyzed and synthesized research. The underlying assumption is that, "medical information will lead, without controversy or politics, to better clinical decision making, better medical care, and better health policy" (Rodwin, 2001, p. 442). The term evidence-based medicine is often used interchangeably with clinical practice guidelines (CPGs). However, it is important to discern the differences. CPGs can be, but are not necessarily, based on evidence-based medicine. In addition, evidence-based medicine can present in forms other than clinical practice guidelines. For example, peer-reviewed journal articles, unpublished studies and expert testimony can all be based on principles of EBM (Rosoff, 2001).

The EBM process essentially involves five steps: (a) defining the question or problem, (b) searching for the evidence, (c) critically appraising the literature, (d) applying the results, and (e) auditing the outcome (Meakins, 2002). Defining the question can be at the level of the specific patient, but is usually centered on a specific problem. However, how the question is defined will impact the next step: searching for the literature. The question must be specific enough in order to gather only the most relevant research studies, without risking being too narrow in scope. Once the literature is obtained, there can, however, be a tendency only to review those studies whose outcomes agree with one's personal perspective. It is essential that each article retrieved be critically appraised in relation to specific criteria such as design, structure, organization, follow-up, analysis and results. The research study's level of evidence is then adduced based on this appraisal. This process claims to prevent personal bias from entering into the treatment decision when applying the results to a particular patient's case. And finally, auditing the outcome involves evaluating the patient's response to treatment and comparing it to the desired or expected result (Meakins, 2002).

The reader will soon recognize that the EBM concepts of critical appraisal and level of evidence mirror the components of health technology assessment; therefore, since these strategies will be used in this research, they will be described in further detail in the Methodology chapter. However, at this point, what will be discussed are the underlying principles for assembling, evaluating, and interpreting medical research. Murow & Lohr (2001) identified nine principles inherent to the many different frameworks that exist for handling this scientific process:

1. A priori explicit statements of questions being addressed.
3. Systematic sorting of relevant from irrelevant research using preset explicit selection criteria.
4. Systematic critique of the validity of individual pieces of medical research based on the quality of the research methodology.
5. Critique of the generalizability of pieces of research based on characteristics of participants involved in research studies and characteristics of the agents or strategies tested in the research.

6. Integration of bodies of evidence based on sources of evidence, research design, directions and magnitudes of clinical outcomes, coherence, and precision.

7. Extrapolation of research findings to particular situations based on preset criteria.

8. Continual updating and integrating of evidence (perpetual revision).

9. Open attribution and statement of conflict of interest by those who do research synthesis.

The research question being proposed will guide the selection of relevant research evidence and the procedures used to critically evaluate and interpret the results. For example, for an inquiry into the efficacy of a treatment, controlled trials would be the evidence of choice. However, for questions pertaining to prognosis, cohort studies may also be acceptable. Each type of study design has a list of specific criteria to evaluate the study against. For controlled trials it would be important to assess factors such as unbiased outcome assessments, concealed randomization procedures and withdrawal rates. For cohort studies, criteria such as representativeness of patient sample, length of follow-up period, and research sponsorship would be important. These criteria will become more detailed and clear when further explained in the Methodology section.

The interpretation of medical research again parallels processes used by health technology assessment researchers. Essentially, practitioners of evidence-based medicine must make a decision in regard to the quality and strength of the evidence at three levels: (a) the quality and applicability of individual studies, (b) the strength and applicability of the body of evidence, and (c) the intensity of recommendations. The quality and applicability refers to the design and conduct of the relevant studies and whether they apply to specific or general populations. Strength and applicability of the body of evidence relates to the weight placed on the collection of research, while the intensity of recommendations signifies the strength of the evidence based on guidelines that may be developed (Mulrow & Lohr, 2001).

At the centre of the evidence-based medicine movement is the Cochrane Collaboration. Cochrane is an independent non-profit organization dedicated to promoting evidence-based medicine through the synthesis of systematically reviewed and critically appraised research. Cochrane was established in 1993 by a group of doctors from the United States, Canada, and the United Kingdom. Its first centre was located in Oxford and was predicated on the principles of the Scottish epidemiologist Archie Cochrane. Over the years the collaboration has expanded to over 11,000 individuals located in more than 90 countries. The Cochrane Collaboration’s web site (www.cochrane.org) lists its guiding principles as:
2. Building on the enthusiasm of individuals.
3. Avoiding duplication.
4. Minimizing bias.
5. Keeping up to date.
6. Striving for relevance.
7. Promoting access.
8. Ensuring quality.
9. Continuity.
10. Enabling wide participation.

The Cochrane has also been the target of much criticism since its inception. To begin with, while Kristiansen and Mooney (2004) believe that, “EBM and the Cochrane Collaboration have achieved a lot, based . . . on some very reasonable ideas” (p. 4), their concerns centered on whether all aspects of EBM should be uncritically accepted. The first issue that they raised was in regard to the expanding scope of the Collaboration. Initially, the Cochrane’s focus was limited to clinical issues, however, much to Sackett’s chagrin, their capacity has extended into other areas of health care such as nursing, health policy, social policy and education. Today, EBM combines approaches to improve the quality of clinical care as well as methods to control health care costs. This has resulted in proponents of the conventional EBM model to conceptually divide evidenced-based medicine into two factions: “Bedside EBM” (or the “real EBM”) and its newer incarnation: “Regulatory EMB” (Dickenson & Vineis, 2002).

The regulatory arm of EBM has been influential in informing policy, commissioning resources and guiding clinical practice. Yet, critics will argue that it is another coercive tool for rationing and auditing health care (Rogers, 2002). One concern is that policy-makers will cite the excuse of lack of evidence to justify setting limits; while another is that the use of clinical practice guidelines may limit access to potentially beneficial treatments (Norheim, 2002). This barrier to access could occur when patients are required to meet specific thresholds for treatment. While it is recognized that evidence obviously plays an important role in limit-setting decisions, Norheim (2002) contends that it is only one of four criteria that should be considered within the broader context of ethical and political concerns: (p. 312)

1. The severity of disease if untreated or treated by standard care.
2. The effectiveness of the new technology.
3. The cost-effectiveness of the new technology.
4. The quality of evidence on the above three points.
The reader will note that the above four criteria are once again similar to the epidemiological and effectiveness contexts of the previously described BCOHTA Comprehensive HTA Framework.

Williams (1997), an economist and general proponent of EBM, defended the regulatory role that EBM has assumed by declaring the importance of comprehensive quantification of health benefits in order to ensure greater accountability. One such measure is the quality adjusted life year (QALY). QALYs do not describe individual symptoms or disabilities. Instead, they are a measure of the overall social value (utility) that a patient places on their state of health. The health rating is a single number on a continuum from death (0) to full health (1.0). This number is then multiplied by the number of years the treatment provides the health state (Sackett et al., 1991). For example, if research showed that living with angina pectoris had a quality score of 0.7, and randomized controlled trials of coronary artery bypass surgery extended life expectancy for another 8 years, then the patient would gain 5.6 QALYs. These measures can then be plugged into economic analyses to determine costs associated with either providing or withholding a treatment. This figure is called the “cost per QALY gained”. Sackett et al. stressed that the purpose of economic analyses should not be on the actual dollar amount attached to a treatment option but instead QALYs should focus on what choices would provide the greatest benefit to patients and the public along with the opportunity costs (“the inability to have enough resources left to do something else” [p.212]). However, an ongoing criticism of QALYs is that interventions that may have very small benefits to many are favoured over treatments that may provide large benefits to the few if total utility is lower in the second case; while another concern is that the same weight is applied to all, irrespective of the patient’s age or current health status (Biller-Andorno, et al., 2002).

Williams (1997) opposed the lopsidedness of EBM’s focus on effectiveness, arguing that cost issues should not be discussed separately; instead the two concepts should be systematically linked and analyzed, “since costs represent health gains that have been denied to others” (p. 120). However, he was sceptical as to the extent that clinical practice guidelines could incorporate cost effectiveness data.

Kristiansen and Mooney (2004) contended that because the core methods of EBM are rooted in medicine, (including biostatistics and epidemiology), and predicated on the paradigm of the randomized controlled trial, the EBM approach may not be adequate to analyze non-clinical issues. Instead, broader theories and methods may be required to answer theoretical questions such as: What is health? How should it be measured? Whose values should be taken into consideration? and, What is meant by evidence? In fact, these authors asserted, “EBM has quickly become a large edifice built on a relatively limited theoretical basis” (p. 5). This position is supported by physician and social scientist, Tor Inge Romøren (as cited in Kristiansen & Mooney, 2004) who also warned against applying EBM’s methods to health services research.
Some of his points included the claims that, (a) it is the responsibility of medicine to be responsible for improving clinical decisions, not health services researchers, (b) that the EBM paradigm is too narrow to be applied to health services, and (c) that EBM is based on rationalism, which does not fit the scheme of health care.

Another criticism was that, although Sackett (1997) claimed EBM was the integration of best research evidence along with clinical expertise and patient values, it did not offer methods to improve clinical skills. Instead, Kristiansen and Mooney (2004) argued that the only contribution that EBM makes to improve clinical expertise is by helping clinicians interpret clinical data.

In the pursuit of "best evidence" some feel that the Cochrane overlooks the limitations of randomized controlled trials (RCTs) and meta-analyses that were acknowledged by Archie Cochrane himself. Kristiansen and Mooney (2004) maintained that RCTs are regarded as superior to other research methodologies based purely on dogma, not scientific proof. "Archie Cochrane, we and many others believe that the design of RCTs is superior in many instances. But belief is not proof or evidence" (p. 10). They reasoned that since there does not exist an external "gold standard" upon which to compare an RCT, there is no empirical evidence to substantiate its superiority claim. Another limitation that they pointed out was that the external validity of an RCT is generally accepted without evidence; therefore results can be extrapolated to inappropriate patient populations. They also argued that there was a potential to bias treatments towards pharmaceutical approaches as the efficacy of a drug is easier to prove with an RCT than interventions such as lifestyle changes. And, in regard to their criticisms of meta-analyses, Kristiansen and Mooney's argument was that, depending on which set of quality criteria were applied against the recovered studies, one could end up with different sets of trials to include in the analysis, and once the analysis began, due to the heterogeneity of research, it would become very difficult to compare like studies (even with the use of statistical tests to account for these differences). In addition, they contended that there was a risk that relevant studies could be omitted from the analysis because they were not identified in the literature search (e.g., unpublished studies or research published in another language).

Power shifts in the health care system have also been identified as another outcome of the advent of evidence-based medicine (Biller-Andorno, et al., 2002). As cost effectiveness analyses and clinical practice guidelines assume a greater importance in resource allocation decisions, health policy-makers, insurance companies and even the pharmaceutical industry leaders may gain a stronghold over a particular treatment.

Another identified issue with evidence-based medicine pertained to the unequal distribution of acceptable evidence. Randomized controlled trials are abundant for pharmacological treatments, emergency medicine and general organic problems, however, there is a paucity of quality research in fields such as neonatology, psychology, and psychiatry.
Opponents decry that EBM challenges the principles of distributive justice because certain branches of medicine become disadvantaged. Dickenson and Vineis (2002) claimed this was due to the difficulty in utilizing a RCT approach in certain conditions, and also the lack of scientific interest and funding opportunities. Although the pursuit of best evidence is recognized as a noble goal, Kasenbrood (as cited in Dickenson & Vineis, 2002) maintained that in reality, in only about twenty percent of the cases is evidence from RCTs available. This could result in treatments being advocated not necessarily because they are better, but because there is more evidence on them.

Discourse also surrounds the various systems of “level of evidence” applied against included studies to grade the evidence to discern the best treatment option. Kristiansen and Mooney (2004) maintained that to date there is no empirically proven “best” method for grading evidence and that these systems are also based on dogma, not on scientific proof. Level of evidence schemes typically place meta-analyses and RCTs on the top rung of the hierarchical ladder; however, Thelle (2004) argued that there is no scientific proof that meta-analyses provide less biased results than cohort studies. As a result of the limitations of RCTs, if improperly conducted, they can produce more bias than some of the other study designs. The concept of level of evidence is also sometimes confused with “level of recommendation”. Here, Kristiansen and Mooney are adamant that the decision whether or not to implement a specific intervention should not be answered by the scientists or clinicians but by patients or policymakers. They argue that the level of evidence is only one component of the decision making process and that value judgments also come into play with the final decision.

While policy-makers, health care managers and payers generally embraced EBM, it was met with some scepticism from within the medical profession. One of the chief complaints of EBM was the physician’s loss of autonomy. Historically, doctors were granted great latitude in clinical decision making. However, with the advent of evidence-based clinical practice guidelines imposed upon the profession by health insurers in an attempt to curb expenditures, some physicians viewed it as being forced to conform to a “cookbook” approach to medicine. However, this perspective was disputed by Sackett (1996) who claimed: “because [EBM] requires a bottom up approach that integrates the best external evidence with individual clinical expertise and patients’ choice, it cannot result in slavish, cookbook approaches to individual patient care” (p. 71).

Another concern expressed by doctors was the perceived increased risk of liability if an evidence-based treatment approach was not followed. As it stands now, “[t]here is sufficient evidence to suggest that most clinicians’ practices do not reflect the principles of evidence-based medicine but rather are based upon tradition, their most recent experience, what they learned years ago in medical school, or what they have heard from their friends” (Eisenberg, 2001, p. 369). However, it was argued that a strict adherence to the tenets of evidence-based
medicine might actually compromise a physician's duty of care to the individual patient (Dickenson & Vineis, 2002; Rogers, 2002). This assertion was based on the inherent population perspective of EBM and the "implicit conflict between the ethics of individual patient care and the ethics of population health" (Rogers, 2002, p. 278). Randomized controlled trials (the foundation of EBM) are predicated on population statistics – an "average" estimate of effects. This can make it difficult to extrapolate results to a particular patient.

While one perspective views EBM as a vehicle to empower patients, providing them with the additional information to make a better decision, others have stated that clinical practice guidelines may in fact reduce patient choice by eliminating other treatment options from the doctor-patient consultation (Dickenson & Vineis, 2002). The ethical question would be whether a physician would be obligated to inform their patient of all the options, even if they were not evidence-based. Some also argued that the recommendations of EBM threaten the return of paternalism, "in that they are the results of experts issuing authoritative instructions as to how to act for the good of the patient, leaving little room for patients to make autonomous choices" (Rogers, 2002, p. 284).

Another critique of EBM was that while systematic reviews and clinical practice guidelines involve judgments about the effectiveness of treatments, these judgments are typically made in isolation of patients' values on health care. "[T]oo often the presumption within EBM is that all that is relevant to the care of the patient is the health outcome" (Mooney, 2004, p. 63). However, it is recognized that health is a complex construct whose definition will vary depending upon the context and culture. Mooney advocated for public participation in determining, "who it is they would want to be the guardians of the values on which health care should be built" (p. 62). For example, in the Netherlands attempts are being made to include patient preferences along side the requisite clinical epidemiological data, pathophysiological knowledge, and clinical experience (Dickenson & Vineis, 2002).

Finally, critics maintained that there was no convincing evidence to support the theory that EBM improves health (Dickenson & Vineis, 2002; Kristiansen & Mooney, 2004; Rasmussen, 2004). Based on the tenets of EBM, convincing evidence would have to take the form of a randomized controlled trial that compared two groups of patients: those that received EBM and those that did not. Obviously, such a study would not be feasible, as one would not be able to isolate patients from receiving some form of evidence-based care. Rogers (2002) claimed that difficulties arise in conceptually defining what is meant by "proof of evidence" since effectiveness is a comparative concept and is based on balancing factors such as benefits vs. harms, the seriousness of the condition, consequences of non-treatment, and the availability of other treatments. She emphasized that these factors must be evaluated alongside the available research evidence. However, often these evaluations are not transparent and as such are
laced with value judgments. This negated the seemingly objective approach of EBM. There were others that claimed that EBM might even serve to increase health care costs (Williams, 1997).

Some suggested that in situations where RCTs are not feasible, the next best level of evidence should be sought. Their stance being that, "evidence-based medicine does not advocate a rejection of all innovations in the absence of definitive evidence. When definitive evidence is not available, one must fall back on weaker evidence . . . and on biologic rationale" (American Medical Association, 1992, p. 11).

**Randomized Controlled Trials (RCTs)**

RCTs provide the strongest evidence for "cause and effect" between the "exposure" and the health outcome. This is due to the inherent properties of randomization, control, and manipulation that provide the necessary and sufficient conditions for definitive research. **Randomization** pertains to the allocation of treatment. With randomization, each subject has an equal chance of being in either treatment arm. It also ensures that the two groups are equal in respect to both known and unknown confounders. **Control** is a more global concept where the investigator is able to manipulate, randomize, and design the experimental protocol. It also includes the use of a comparison, control group (subjects that do not receive the treatment). And finally, **manipulation** refers to the investigator's ability to apply the experimental treatment.

Unfortunately, merely conducting an RCT in no way guarantees quality evidence. Each study should be evaluated for threats to validity. This includes the evaluation of chance, bias and confounding as possible alternative explanations for observed results. **Chance** can be assessed by an examination of the p-values and confidence intervals. P-value interpretation is often limited, however, because of its dependence on sample size and the size of the effect. Because of this, the confidence interval is usually more accurate with a narrow interval indicating more precision and a wide interval representing more chance variation. A definitive study would have a significant p-value and a narrow confidence interval. **Bias** is, "any systematic error . . . that results in an incorrect estimate of the association between exposure and risk of disease" (Hennekens & Buring, 1987, p. 272). Conditions that support valid results would be evidence that the researchers attempted to control biases in design, conduct, and analysis. Studies should be examined for sources of potential biases (information, selection, lead time, and incidence prevalence biases), the direction of any likely effect and how large the bias may be. And finally, **confounding** is, "the mixing of the effect of the exposure under study on the disease with that of a third factor" (p. 287). This factor is associated with the exposure and is a risk factor for the disease. Evidence must show that actions were taken by the investigator to control for potential confounding. This can include matching, randomization, and the analytical technique of stratification.
Alternate Study Designs

As previously mentioned, RCTs are not always feasible, economical, or ethical to conduct. Therefore, alternative study designs may need to be examined. In a case-control study, subjects are selected based on their current disease status (+ disease, or - disease) and compared in regard to the odds of past exposure. A benefit of this design is that it does not require follow-up measurements, thereby making it quick, efficient, and especially suitable to the investigation of diseases with long latent periods. However, case-control is prone to threats to validity. In cohort studies subjects are assigned to groups based upon their exposure status and then followed forward to an outcome. One advantage is that it is easier to make a temporal association between the exposure and outcome because subjects were “disease free” at the outset. However, a major drawback of cohort studies is that they are time-consuming, expensive and their results could be impacted by loss to follow-up (research subjects that can not be contacted for follow-up assessments, due to reasons such as early withdrawal, a change in geographical location, or refusal).

Clinical Practice Guidelines (CPGs)

CPGs are official recommendations for best practices that are developed by multidisciplinary teams utilizing systematic, evidence-based methodologies. The specific approach for developing CPGs can differ depending on: (a) the criteria used for identifying, appraising and ranking relevant research; (b) the selection of other models which incorporate indirect evidence, opinion and experience; (c) strategies for balancing harms, costs, and values; and (d) consideration of research sponsorship (Mulrow & Lohr, 2001). One of the most widely recognized approaches to developing CPGs was developed by the Agency for Healthcare Research and Quality (AHRQ, 2002) and consists of the following main elements:

1. A standardized methodology for screening and systematically evaluating relevant scientific literature.
2. Use of multidisciplinary consensus panels.
3. Recommendations based on panel opinion when there is a paucity of scientific evidence.
4. Details provided regarding methods used, evidence examined, and the panel’s rationale for each guideline developed (including strength of evidence).

Included in the AHRQ development process are an assessment of both the quality and the clinical applicability of the scientific evidence. Quality assessment is based on a judgment of how well a study has been designed to control for bias, while clinical applicability is the degree to which a study’s results would occur in a particular clinical situation. Once a study meets the quality and applicability criteria, the data are abstracted into summary evidence tables to ease the comparison process in order to assign an evidence rating score. Once the final guideline recommendation is developed, it is then evaluated and rated based on its strength of
evidence. This is a classification system whereby the amount of evidence, its general quality, and its clinical application are evaluated. Mulrow and Lohr (2001) identified four key concepts to developing defensible guidelines:

1. The development process should be open, documented, and reproducible.
2. The resulting product should be of use to both patients and clinicians.
3. The guideline should promote “appropriateness” of services (health benefits outweigh potential harms).
4. The guideline should relate to specific, clearly defined clinical issues.

An underlying assumption of clinical practice guidelines is that, “attempts to develop practice guidelines based on evidence and science [are] a way to get physicians to rise above their parochial views and self-interest” (Rodwin, 2001, p. 442). However, as was the case with consensus development, the mere existence of CPGs does not necessarily result in change in a physician’s practice. It is recognized that financial interests and personal biases can get in the way. Often times these are perpetuated by administrative policies dictated by governments (e.g., fee-for-service) and hindered by institutional controls (e.g., operating room space). Each of these conditions impacts the choices physicians can make. Other criticisms included the artificiality of guidelines by attempting to break down medical work into a series of successive demarcated steps (ter Meulen & Dickenson, 2002). Rodwin (2001) proffered some suggestions for promoting the practice of medicine based on evidence. He proposed there should be rewards for physicians that elect to follow practice guidelines that are evidence-based and that institutions should implement processes that, “promote physician oversight and accountability” (p. 443).

**Standards**

Standards differ from CPGs in that, while CPGs are flexible and customized to each patient, standards are stringent, medically necessary practices that must be followed. Standards raise the evidentiary bar because they require additional priority setting based on cost and value judgments (Mulrow & Lohr, 2001). Therefore, standards are understood to be, “enforceable by Courts in malpractice cases and other legal disputes” (Havighurst, et al., 2001, p. 201). A practitioner that fails to implement a guideline may not be held liable while a physician that does not carry out a standard could be deemed negligent.
Conceptual Framework

While Auton has been previously examined through various lenses (Freeman & Mayerson, 2003; Greschner & Lewis, 2003; Syrett, 2005), this thesis situates the case within a comprehensive health technology assessment framework (Kazanjian, 2004), which facilitates an update of the effectiveness evidence on Lovaas therapy. In addition, the legal context of the social dimension of the HTA framework is further expanded to permit a conceptual analysis of the role of evidence within the domains of law, health policy and health care. Appendix A depicts the conceptualization of this process.

More specifically, for the analysis of the legal context, I developed a conceptual framework based on the findings from the IOM/AHRQ workshop (Eisenberg, 2001). As previously described, these discussions culminated in the identification of six opposing conceptualizations of evidence, in addition to proffering preliminary conceptual stances on the concept of evidence for the domains of law and health care. Figure 2.1 illustrates my conceptualization of the common context in which “evidentiary actors” (change agents, gatekeepers and decision-makers) operate within; the figure also depicts the targeted level of application of the evidence.

**Figure 2.1 The Context, Actors and Level of Application of the Evidentiary Process**

From this preliminary conceptualization, I developed distinct scientific evidence pathways for each domain. The resultant framework (Appendix B) identifies seven categories of variables that influence each sector’s conceptualization of evidence: (a) perspective of evidence, (b) level of evidence, (c) rules of evidence, (d) gatekeepers of evidence, (e) decision-makers of evidence, (f) change agents of evidence, and (g) level of application of evidence.
Perspective of evidence is a contextual variable that corresponds to the IOM/AHRQ's population probabilities vs. individual causation category. It refers to the level of evidence utilized for decision making (e.g., population-based findings vs. evidence specific to an individual within a particular context). Level of evidence, the second contextual variable, relates to whether the evidence originates from primary research (e.g., RCTs) or secondary synthesized designs (e.g., systematic reviews). Rules of evidence is the third contextual category of variables that guides each sector's interaction with evidence. The next three categories of variables – Gatekeepers, Decision-makers, and Change Agents – were identified as the key "evidentiary actors" in the process of seeking, understanding, and applying evidence within their respective professional/sectoral context. The level of application of evidence pertains to the target audience of the specific decision (based on the evidence). The final variable is pre hoc or post hoc utilization of evidence and is graphically depicted in the framework by positioning the "outcome" variable (represented by a circle with a cross through it) along each domain's pathway either before or after the evidentiary process, based on their theoretical position. And finally, my framework illustrates that each domain shares a common goal of generalizing the evidence for the betterment of society.

This chapter has presented the multi-disciplinary perspectives on evidence that have culminated in my conceptual framework for analysis. This will be the lens in which the methodologies of the following chapter will be applied.
CHAPTER 3: RESEARCH DESIGN AND METHODOLOGIES

The first objective of this thesis is an examination of two legal decisions on the provision of government-funded autism treatment for specific children in British Columbia (Auton and Anderson) in order to discern how evidence is sought, understood and applied within legal, health policy and health care domains. The methodology utilized was a grounded theory approach to the analysis of interview transcripts and relevant legal documents. This chapter provides an in-depth description of grounded theory along with details of the specific methods utilized to collect, code and analyze the data. In addition, this chapter offers an introduction to the qualitative research software program - QSR N6 - that was used to assist in the interpretation of the interview data and legal documents.

The second objective of this research is an appraisal of the effectiveness evidence on Lovaas Autism Treatment. This was conducted through the methodologies of systematic review and critical appraisal. This chapter not only presents the background information on these methodologies but also outlines the specific approaches used for the literature search, study selection, quality assessment, data extraction, and descriptive data synthesis.
Grounded Theory: The Legal Context

Grounded Theory Methodology

Grounded theory is a general qualitative research methodology used for developing theory; in essence, it is a way of thinking about and conceptualizing data. The resultant theoretical framework has the hallmark of being grounded in data that have been systematically gathered and analyzed (Strauss & Corbin, 1994). The goal of grounded theory is, "to generate a theory that accounts for a pattern of behaviour which is relevant and problematic for those involved" (Strauss, 1987, p.5). Theory building involves the steps of conceptualizing, defining categories, refining categories in terms of their properties and dimensions, and relating categories through hypotheses or statements of relationships.

Conceptualizing involves grouping like items according to defined properties and then labelling the concept. Once categorized, defining categories entails discovering the specific properties that describe the category and then determining how they vary dimensionally along those properties (refining categories). These procedures enhance the data in order to view patterns, which are the initial pieces of the grounded theoretical puzzle (Strauss & Corbin, 1998).

Grounded theory is rooted in the epistemological tradition of Interpretivism and the philosophy that, "social science [should] address the meaningful character of human life" (Travers, 2004, p. 8). Interpretivism highlights, "the dynamic, constructed and evolving nature of social reality" (p. 8), by understanding social reality through the eyes of those being studied. Within the tradition of Interpretivism lies the philosophy of symbolic interactionism. Although the American sociologist Herbert Blumer first coined this term in the 1950s, the approach of naturalistic study actually began in 1913. Its theoretical underpinnings originated in the early writings of the pragmatist philosopher Charles Horton Cooley and the social psychologist George Herbert Mead, while its fieldwork evolved from the Sociology Department of the University of Chicago between 1913 and 1933 (under the direction of Robert Park). The assumptions held by symbolic interactionists are that social meanings do not reside solely in the minds of individuals but are shared by members of a society or by particular social groups; meaning is therefore conceptualized as an inter-subjective rather than a subjective characteristic. The key point of symbolic interactionism is that, "individuals are influenced by other people, but . . . they are also active in interpreting, and responding to, the people and objects they encounter in the world" (Travers, 2004, p. 42).

Since these early origins of symbolic interactionism, over three generations of academics have diversified this discipline into various unique approaches - grounded theory being just one of them. American sociologists Anselm Strauss and Barney Glaser were co-founders of the grounded theory sub-tradition. In the mid 1960s Strauss and Glaser were hired by the University of California San Francisco to assist nursing students with their research. Their first project - studying dying patients in hospitals - led to the development of this new interactionist method,
which proved to be a more rigorous, scientific version of symbolic interactionism (Glaser & Strauss, 1967). Earlier symbolic interactionists such as Blumer had felt that human beings could not be studied effectively in a quantitative way. However, Glaser and Strauss deviated significantly from this view in believing that by using grounded theory researchers were able to produce theoretical propositions that were testable, verifiable and capable of predicting future events. This conviction was supported by detailed methods that utilized a set of replicable procedures. The key feature of Strauss and Glaser’s methodology is that theory evolves during the actual research through a continuous interplay between data analysis and collection. It goes beyond the goal of other qualitative methods that produce rich descriptions of social interactions. Theory development is therefore the key feature of grounded theory methodology.

Glaser and Strauss (1967) defined theory as the following: 

Theory in sociology is a strategy for handling data in research, providing modes of conceptualizations for describing and explaining. The theory should provide clear enough categories and hypotheses so that crucial ones can be verified in present and future research; they must be clear enough to be readily operationalized in quantitative studies when these are appropriate. (p. 3).

What sets grounded theory apart from other qualitative approaches is its ability to develop conceptually dense theories containing many conceptual relationships. These relationships are initially stated as propositions, but they are strengthened through continual research. Grounded theorists are interested in patterns of action and interaction between and among social actors. In particular, they are interested in discovering process, “not in the sense of stages or phases, but of reciprocal changes in patterns of action/interaction and in relationship with changes of conditions either internal or external to the process itself” (Strauss & Corbin, 1994, p. 278). Strauss & Corbin asserted that since theories developed through this approach have the ability to specify consequences and their related conditions, they can also claim predictability in as much as if similar conditions apply somewhere, then similar consequences should occur. However, grounded theories are described as “fluid” in that they engage many actors and conditions and involve temporality and process. For this reason grounded theories are described as, “not just another set of phrases; rather, they are systematic statements of plausible relationships” (p. 279). Strauss and Corbin (1994) address the debate over the relationship of theory to reality and truth:

Our position is that truth is enacted: theories are interpretations made from given perspectives as adopted or researched by researchers. To say that a given theory is an interpretation – and therefore fallible – is not at all to deny that judgments can be made about the soundness or probable usefulness of it . . . theories are embedded “in history” – historical epochs, eras, and moments are to be taken into account in the creation, judgment, revision, and reformulation of theories. (p. 279)
Substantive theory is grounded in research on one specific substantive area; these theories can provide the initial direction for more formal or "general" theories since relevant categories and properties may have already been developed. While formal theory can evolve directly from the data, most often it is created through analysis and assimilation of substantive theories.

**Assumptions of Grounded Theory**

The following outlines the major assumptions of grounded theory:

1. The human status of actors is studied.
2. Actors have perspectives on and interpretations of their own and other actors' actions.
3. Social meanings do not reside solely in the minds of individuals but are shared by members of a society or by particular social groups.
4. Meaning is conceptualized as an inter-subjective rather than a subjective characteristic.
5. Researchers are able to produce theoretical propositions that are testable, verifiable and capable of predicting future events.

**Critiques of Grounded Theory**

Strauss and Glaser began to differ on both epistemological and methodological aspects of grounded theory when in 1990 Strauss joined forces with nurse researcher Juliet Corbin to publish the first edition of *Basics of Qualitative Research*. Glaser's perspective had been from a qualitative paradigm where he viewed grounded theory in a more laissez-faire light. For instance, he saw research being guided strictly by the informants and their socially constructed realities. His view being that their world should be allowed to emerge naturally from the data without a lot of restrictive processes imposed by the researcher. He opposed the excess of rules and procedures that Strauss had developed for working the data fearing that it forced the data into a pre-defined paradigm. Glaser's (1992) *Basics of Grounded Theory Analysis: emergence vs. forcing* was a direct attack on Strauss and Corbin's 1990 book. Glaser sought to correct what he perceived as errors made by Strauss and Corbin by writing a concise comparison between their two philosophies of grounded theory. His argument was that Strauss' version was not grounded theory; instead, he described it as full conceptual description.

Strauss, on the other hand, had a more scientific quantitative bent towards grounded theory, advocating responsibility for retaining the canons of good science such as replicability, generalizability, precision, significance, and verification. In addition, while Strauss tackled research issues of relevance to individuals, his perspective was more focused on a description of the larger cultural scene. Glaser and Strauss also differed in regard to their conceptualization of the research question. Strauss contended that research problems could come from sources such as the technical literature, personal or professional experiences, funded projects or even topics assigned by a professor to a student. Glaser on the other hand felt that the research
problem itself was discovered during the process of open coding, theoretical sampling, and constant comparison. His position being that the grounded theorist need only begin with a curiosity of what was going on in relation to a specific issue and how it was handled. Glaser also criticized Strauss’ position on theory verification stating that it fell outside the realm of grounded theory (Babchuk, 1997).

Glaser and Strauss also have opposing views on the role of literature and methods of theorizing (Hunter, et al., 2005). Glaser believes literature should initially be used to gain an overarching perspective of the research problem and then again later on to confirm the theory. This is achieved firstly through inductive processes followed by deductive methods once the theory has emerged. On the other hand, Strauss’ method utilizes the literature to identify phenomenon and to analyze data through deductive means. Theorizing through inductive versus deductive processes is a contentious issue amongst social scientists. Induction involves “deriving concepts, their properties, and dimensions from data” (Strauss & Corbin, p. 22), while deduction entails, “hypothesizing about the relationships between concepts, the relationships are also derived from data, but data that have been abstracted by the analyst from the raw data” (p. 22). Strauss and Corbin are quick to point out that while many qualitative researchers anguish over the “right way” of collecting data and validating hypotheses, their position is that there are many ways of achieving the same results and “anytime that a researcher derives hypotheses from data, because it involves interpretation, we consider that to be a deductive process” (p. 22).

One systematic review of grounded theory research in adult education (Babchuk, 1997) discovered that many researchers viewed grounded theory more as an umbrella term incorporating many various diverse methods. This ranged from scholars using but one component of grounded theory (e.g., constant comparison); academics who chose methods that best “fit” their questions; studies that utilized mixed methodologies; right up to those researchers that attempted to follow the exact procedures of this tradition. The ongoing debate is whether grounded theory’s flexibility of application is a methodological strength or a weakness (Strauss and Corbin themselves do not advocate rigid adherence to their procedures).

And finally, Strauss and Corbin (1994) point out that some researchers who claim to be utilizing the grounded theory methodology in actual fact fail to incorporate many important aspects of this approach. For example, they may focus so intently on coding but fail to undertake theoretical coding (more to follow on this topic). And, even if theoretical coding is carried out, unless undertaken alongside constant comparison, theories may not be so conceptually rich as they could be. Or, some researchers may not be interested in developing theory and therefore use this methodology inappropriately.
Grounded Theory Methods

Grounded theory utilizes a detailed, systematic process for simultaneously collecting, coding, and analyzing data for the purpose of studying a phenomenon. *Phenomenon* is a term that answers the researcher's question – "What is going on here?" Strauss and Corbin (1998) state that:

In looking for phenomena, we are looking for repeated patterns of happenings, events, or actions/interactions that represent what people do or say, alone or together, in response to the problems and situations in which they find themselves. (p. 130)

Within the context of this thesis, categories that represented the phenomena of how evidence is sought, understood and applied by the three sectors of law, health policy and health care were initially identified in the literature and subsequently incorporated into my conceptual framework. Themes that emerged in the data were compared to these categories and contrasted across each sector. New themes were assigned their own categories. The methods employed in this thesis were based on Strauss and Corbin's (1998) approach to grounded theory.

Data Collection Methods

**Source.** The data sources for this thesis included publicly accessible legal documents (*Reasons for Judgment*) from the Auton and Anderson cases. These documents were retrieved through computer searches of provincial and federal Court databases. The second data source was interviews with key informants from legal, health policy and health care disciplines.

**Participants.** A small representative sample of six participants were interviewed. This included members of the legal (L), health policy (HP) and health care (HC) sectors (two from each sector). Six participants were chosen as per Strauss and Corbin's suggestion that, "Microscopic coding of 10 good interviews or observations can provide the skeleton of a theoretical structure" (1998, p. 281). Therefore, since the objective of the thesis was not theory development – but instead, conceptual ordering as an initial first step – six participants seemed sufficient. Although medical doctors were approached, none responded within the time frame; therefore psychologists were interviewed. In addition, one Supreme Court judge was also approached and declined. Some of these research participants were found by searching government databases and through referrals from experts in the field of autism. Each participant was initially approached with a Letter of Introduction (Appendix C) explaining the purpose of the study. They were also sent the UBC Behavioural Research Ethics Board-approved consent form (Appendix D) for information purposes only; if they agreed to participate the consent was signed in person on the day of the interview (with a copy retained by the participant). One participant also requested a copy of the interview tape.

**Interview methods.** The interviews were conducted at the participants' place of work between May and November 2006. Each interview was limited to one hour. Written permission to tape-record the interview for future transcription and analysis was granted by all subjects.
Interview questions were derived from the conceptual framework of evidence (Appendix B) with a corresponding question developed for each respective sector. The role of the researcher was to state the question, and if clarification was required, to explain the corresponding conceptual theme behind it. Additional questions were also interjected when elaboration on a concept was deemed necessary.

**Microanalysis**

Grounded theory involves a focused microanalysis of the data in order to generate (or "code") the initial categories. For this task the "unit of analysis" can be a word, line, sentence or paragraph; for this thesis the unit of analysis was a sentence. Microanalysis encourages the researcher to consider all ranges of plausibility in the data. Questions asked were not purely descriptive but incorporated conceptual questions that, "uncovered the properties, dimensions, conditions, and consequences such as who, when, what, how and why" (Strauss & Corbin, 1998, p. 66). This permits data to be broken apart and reconstructed in an interpretive scheme.

**Coding**

Grounded theory is based on the concept-indicator model that designates conceptual codes for a set of empirical indicators (Strauss, 1987). **Empirical indicators** are actual data in the forms of behavioural actions and events; these can either be observed (e.g., in the field) or, as is the case with this thesis, described in documents and interviews. These data are **indicators** of a **concept** identified by the researcher.

---

**Figure 3.1 Indicator-Concept Model**

![Figure 3.1 Indicator-Concept Model](image)

CONCEPT: A class of actions/events

Indicators compared to concept

INDICATORS: actions/events (compared to each other)

CODES

(Source: Adapted from Strauss, 1987, p. 25)

Coding is an essential feature of grounded theory. It consists of a systematic process of, (a) describing data, (b) conceptually ordering data, and, (c) theorizing about data. In addition, coding can be categorized as open, axial, or selective. Open coding is the analytic process whereby initial concepts and their relative properties and dimensions are discovered in the data and organized into categories. Concepts are the building blocks of theories and can be defined through a specific set of properties (characteristics) and dimensions (the range along which characteristics can vary). Categories can also include subcategories that contain concepts pertaining to the main concept (Strauss & Corbin, 1998). Codes can either be socially...
constructed or in vivo. Socially constructed codes are developed by the researcher based on their knowledge of both the literature and the substantive field. They tend to broaden the local meaning of a concept towards a more sociological view. In vivo codes, on the other hand, are verbatim terms used by the actors. They often reflect specific behaviours or processes the actors use to solve their problem (Strauss, 1987).

Axial coding is, "the process of relating categories to their subcategories, termed 'axial' because coding occurs around the axis of a category, linking categories at the level of properties and dimensions" (Strauss & Corbin, 1998, p. 123). The purpose of axial coding is to rebuild the data that were disassembled during open coding. However, since coding in general is a fluid process, axial coding does not follow linearly behind open coding, as often a sense of how categories relate can develop during open coding.

In order to conduct axial coding, knowledge of a few key concepts is integral. First is the term structure. Structure is "the conditional context in which a category (phenomenon) is situated" (Strauss & Corbin, 1998, p. 123); it provides answers as to why certain events occur. The second concept is process, which is "sequences of action/interaction pertaining to a phenomenon as they evolve over time" (p. 123); it provides the reasons behind how persons acted or interacted. By incorporating questions such as why, where, when, how, and with what results, into the analysis, the researcher is able to link structure and process in order to contextualize the phenomenon under study.

Sorting and organizing relationships between events and happenings are facilitated through the use of a paradigm. A paradigm is, "a perspective taken toward data, another analytic stance that helps to systematically gather and order data in such a way that structure and process are integrated" (Strauss & Corbin, 1998, p. 128). A paradigm consists of the following components: (a) conditions, (b) actions/interactions, and (c) consequences. Conditions are a conceptual way of grouping answers discovered by posing questions of the data such as why, where, and when. Next, actions/interactions are the routine or strategic responses, performed by individuals or groups to issues, problems, happenings or events. These are discovered through the proposition of questions such as whom and how. And, finally, consequences are uncovered in the data through questions that pertain to the outcomes or results of actions/interactions by these individuals or groups (or the failure to act/interact).

Selective coding is, "the process of integrating and refining theory" (Strauss & Corbin, 1998, p. 143). At this stage (which does occur linearly after both open and axial coding), major categories are integrated to form a larger theoretical scheme that eventually becomes the theory. Integration begins with the identification of a central or "core" category that explains what the research is all about. A core category should be the "main theme" for the specific problem or concern that the actors are facing and should account for most of the variation in a pattern of behaviour (Strauss, 1987). Provisionally identifying a core category begins during the
open coding phase and proceeds throughout all analytical steps. Certain techniques can aid in discerning the core category including writing a storyline, developing diagrams, and reviewing and sorting memos. Strauss (1987) has outlined six criteria for identifying a core category: (p. 36)

1. It must be central, that is, related to as many other categories and their properties as is possible.
2. The core category must appear frequently in the data.
3. The core category relates easily to other categories.
4. A core category in a substantive study has clear implications for a more general theory.
5. As the details of a core category are worked out analytically, the theory moves forward appreciably.
6. The core category allows for building in the maximum variation to the analysis.

The second stage of selective coding is refining the theory, which consists of, "reviewing the scheme for internal consistency and for gaps in logic, filling in poorly developed categories and trimming excess ones, and validating the scheme" (Strauss & Corbin, 1998, p. 156). Here, the researcher is looking for a consistent, logical flow to the theoretical scheme with a central category that is well defined in regard to its properties and dimensions. These criteria should all be supported by references to data as documented in memos. Upon review of the scheme, if the researcher discovers some "weak" categories, either reviewing memos and/or returning to the field to conduct more focused theoretical sampling to strengthen the properties and dimensions of a category is possible. However, at some point the researcher must determine when theoretical saturation has occurred and stop the data collection process (Theoretical saturation is the point in which no new properties and dimensions emerge from additional data gathered).

The final step in refining the theory involves validation to determine how well the theoretical scheme fits the raw data and whether anything has been left out. This can be accomplished in a couple of different ways. First, a high-level comparative analysis of the scheme against the original raw data can be conducted. Or, the story can be told to the people interviewed/observed in order to elicit their opinions as to its overall "fit" with their reality. With either approach, the goal is to be able to explain the majority of the cases through the theory.

Two other techniques worth mentioning are memos and diagrams. Memos are "written records of analysis" (Strauss & Corbin, 1998, p. 217). They are conceptual (as opposed to descriptive) and contain the products of a researcher's analytical sessions in addition to providing suggestions for directions to take the research. Diagrams, on the other hand, are visual representations of the identified conceptual relationships.
Data Analysis

Data for grounded theory can be very diverse incorporating written texts, videos, interviews, and observations to name but a few. The unique feature of grounded theory methodology is that its processes are not linear — in other words, all data are not coded first and analyzed later. Data collection and analysis are ongoing processes. Analysis can begin as soon as the first couple of interviews are conducted (or documents reviewed). Analytic questions and hypotheses about categories and their relationships should arise during this early analysis and are used to direct subsequent interviews or observations. There is a constant interplay between the researcher and their data while both gathering and analyzing the results. It is recognized that this interplay is not a purely objective process. Instead, Strauss and Corbin (1998) acknowledge that researchers bring their unique disciplinary and research experience into the analysis in an effort to enhance the creative aspects of analysis as opposed to driving it. Researchers’ experience can be invaluable in sensitizing themselves to the potential properties and dimensions in the data. Analysis of the data takes into consideration both the participants’ recollection and interpretation of actual events and actions. Researchers use techniques to abstract, reduce and relate categories and their relationships; this is key to differentiating theoretical coding from mere descriptive coding (theory building vs. description).

Although grounded theory methods were utilized to collect, code, and analyze data, the goal of this research is not the development of theory to explain and predict the phenomenon of the role of evidence; this would have been too lofty an objective based on the emerging state of knowledge on this topic. Instead, what is initially required is basic knowledge and an understanding of how each sector views evidence. This methodology provides a useful approach for analyzing legal discourse through a transparent map of conceptual thinking. Therefore, grounded theory was not employed to its fullest extent. Instead, its lower level of abstraction — conceptual ordering — was utilized to identify and organize preliminary concepts and to begin to define them based on their properties and dimensions. By default, conceptual ordering also includes the lowest order of abstraction — description — which is, “the use of words to convey a mental image of an event, a piece of scenery, a scene, an experience, an emotion, or a sensation; the account related from the perspective of the person doing the depicting” (p. 1).
QSR N6 Qualitative Research Software

The QSR N6 qualitative computing program was selected to aid in the indexing, searching and analysis of the data for this thesis. Once a grounded theory methodology was selected, the search was on for a compatible software program. Initially, N6's predecessor - NUD*IST- was recommended by a colleague and a subsequent literature review revealed not only its "fit" with the descriptive/interpretive or theorizing approaches of grounded theory methodology, and the type of data (textual) that was going to be utilized, but also its utilities supported the methods required of grounded theory research. In fact, N6 was developed on the backbone of grounded theory, which is the predominant approach of most N6 users (Williams, Mason & Renold, 2004).

QSR N6 is a software program designed to manage and explore qualitative data for research projects that utilize coding-based methodologies. N6 is the sixth version that includes as its predecessors N4, N5, NVivo and various series of NUD*IST software (which is an acronym for Non-numerical Unstructured Data Indexing Searching and Theorizing). N6 is essentially a toolkit for coding, analyzing and exploring text documents. The program is generic in that it can incorporate different qualitative methodologies and philosophies and has the capability of importing large amounts of various data, such as transcripts from interviews or focus groups, structured quantitative questionnaires, free text, journals, and documents (Richards, 2002; QSR, 2002).

For grounded theory, N6 supports the coding of categories ("Nodes"), which can be explored, tested, modified, and linked. This is enhanced by the researcher's ability to write memos directly into the program to document each analytical decision along the way. N6 provides the ability to query both the imported text documents and the nodes in order to search for specific passages, words, phrases, or ideas (QSR, 2002).

[Qualitative research] requires sensitivity to meaning and context, accurate access to information and ways of rigorously and carefully exploring themes and discovering and testing patterns. It always requires that researchers record growing understanding in summaries, annotations or memos or field notes. Qualitative research also treats as data the records of ideas about these research events and reflections on them. (Richards, 2002, p. 3)

Since the early 1980s when computers first became popular for aiding qualitative researchers, there has been much debate over the use of Computer Assisted Qualitative Data Analysis Software (CAQDAS). Reservations include the concern that the computer takes control of the data by directing the researcher to handle data in a specific way thereby distancing the researcher. There can also be a risk of substituting coding for analysis and developing large, cumbersome hierarchical coding trees which may preclude effective data analysis.
However, advocates of CAQDAS assert that textual analysis packages ease a researcher's workload, save time and serve to enhance the power of qualitative analysis (Buston, 1997). "The role of the computer remains restricted to an intelligent archiving ("code-and-retrieve") system, the analysis itself is always done by a human interpreter (Kelle, 1997: para. 15.7). Buston (1997) contended that CAQDAS "are without utility in making the jump from ordered data to theories (and will be without such utility until they incorporate sophisticated artificial intelligence)" (para 12.4). However, Buston did acknowledge that programs such as N6 are not neutral tools. There are inherent risks to continue coding data incessantly and to continue the quest for additional data. Their argument was that while the program influences the researcher, this was viewed as advantageous because it allowed for additional time for data analysis. However, as long as the researcher kept the reason for doing things in sight, these pitfalls could be avoided. The capabilities of the software package should not dictate the handling of the data.
Systematic Review and Critical Appraisal: The Effectiveness Context

The comprehensive health technology assessment (HTA) framework developed by the British Columbia Office of Health Technology Assessment (BCOHTA) was the methodological approach used to evaluate the effectiveness of Lovaas Autism Treatment (Kazanjian, 2004; Kazanjian & Pagliccia, 1998). This approach was selected in order to seamlessly update BCOHTA's previous Lovaas review (Bassett, et al., 2000). As previously described, HTA can provide an assessment of the safety, efficacy and effectiveness, as well as the financial, ethical, legal and social implications of a proposed technology. The scope for this study is the effectiveness and legal dimensions of the HTA framework.

Systematic Review

Because of the sheer volume of health care research available, some of which is of questionable quality, systematic reviews are becoming essential components of cost effective and efficient health care decision making. Slowly, they are beginning to replace historical methods such as opinion-based narrative reviews, consensus development statements, and expert opinion. A systematic review is a rigorous approach to collating scientific evidence on health care issues such as treatment, diagnosis, or preventive services. It utilizes a comprehensive, systematic, objective, and transparent process that limits bias in the study search and selection procedures in addition to the data extraction and synthesis phases. Systematic reviews that incorporate critical appraisal also assess the methodological quality and evaluate the overall strength of body of evidence of the studies selected. Although the term systematic review is often used synonymously with meta-analysis, this is not an accurate usage as meta-analysis is a statistical methodology that is used to combine data from different studies into a single summary estimate. Systematic reviews can be conducted in combination with a meta-analysis when there is homogeneity between the studies. Otherwise, systematic reviews can consist of either a systematic literature review on its own or in combination with a critical appraisal process.

Review Question

The review question for the effectiveness assessment portion of this thesis is: "How much improvement in cognitive, functional, intellectual, linguistic and social functioning do autistic children under the age of 6 achieve with Lovaas-type applied behavioural analysis therapy?"

Search Strategy

The systematic literature search strategy utilized a comprehensive approach to obtain the available evidence. In general, the electronic search accessed commercial databases, historic BCOHTA databases, web library catalogues, Internet peer-reviewed sites, Internet search engines, directories and organizations. There was no language restriction placed on the search. More specifically, Medline, EMBASE, CINAHL, Cochrane, and PsycINFO electronic bibliographic databases were initially searched on OVID for the period from January 1998 to December 2004.
with updated searches conducted in December 2005 and again in March 2006 (Appendix E). In addition to the electronic search, a review of reference lists of relevant articles was conducted along with a search for grey literature (literature that is not indexed or distributed publicly) through various extended search methods. In order to ensure comprehensiveness of the search strategy, substantive experts were also identified and sent a list of retrieved studies.

**Study Selection Criteria**

Study selection consisted of a review of relevant primary studies and systematic reviews based on generally accepted selection criteria: (a) population of interest, (b) intervention, (c) outcome measures, and (d) study design (Alderson, Green & Higgins, 2004; Canadian Coordinating Office for Health Technology Assessment (CCOHTA), 2003; NHS, 2001).

<table>
<thead>
<tr>
<th>Selection criteria</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
</table>
| **Population**     | Autistic children < 6 years  
Dagnosis of autism spectrum disorder | All others |
| **Interventions**  | Lovaas-type applied behavioural analysis compared with another intervention | Lack of comparison group/intervention |
| **Outcome measures** | Intellectual functioning  
Language  
Social interaction and play  
Adaptive or self-care skills  
Maladaptive behaviour  
Educational functioning | All others |
| **Study designs**  | Systematic reviews  
Studies utilizing treatment and control groups:  
o Randomized Controlled Trials  
o Cohort studies  
o Case control studies | Lack of a control group:  
o Case reports  
o Case series |

**Rationale for Selected Inclusion/Exclusion Criteria**

**Population.** Consistent with the theories of neurobiological development (Bailey, Aytch, Odom, Symons & Wolery, 1999), and results of early Lovaas studies (Lovaas, Koegel, Simmons & Long, 1973), treatment should be provided during the preschool years. Therefore, primary studies had to stipulate inclusion criteria for autistic children less than 6 years of age. In relation to diagnosis, subjects had to be diagnosed with autism spectrum disorder utilizing DSM-IV criteria. All other pediatric populations were excluded.

**Interventions.** For inclusion in the review, the treatment intervention had to be described with key words such as “intensive applied behavioural analysis”; “Lovaas”; “Lovaas autism treatment”; “early intensive behavioural intervention”; “early intensive behavioural treatment”;
and/or "intensive discrete trial training". No restriction was placed on type of control intervention but the absence of a comparison group/intervention resulted in exclusion of a study from further consideration.

**Outcome measures.** Any form of intellectual, behavioural or language outcomes was sufficient.

**Study designs.** Since previous systematic reviews discovered the lack of randomized controlled trials of Lovaas therapy (Basset, Green & Kazanjian, 2000; ECRI, 2000), non-randomized studies were also assessed provided they utilized a treatment comparison control group. In addition, systematic reviews were also included.

**Study Selection Process**

Study selection was conducted by three researchers in an un-blinded fashion. While some reviewers advocate blinding studies for critical appraisal, the merits of blinding reviewers to elements of a study that could cause bias have not been empirically consistent. For instance, Jadad et al. (1996) showed that blinded assessment produced significantly lower and more consistent scores (indicating a poorer quality of study) than un-blinded assessments, while Moher et al. (2004) produced the opposite results with a small absolute increase in scores of 3.8% in blinded studies. And yet, another study found no difference between the two methods (Clark et al., 1999). The general consensus in the literature revealed that the methodological challenges and financial implications of blinding systematic reviews did not warrant the slight reduction in bias that may be achieved (Alderson et al., 2004; Clark et al., 1999; NHS, 2001).

Therefore, the first reviewer (SB), second reviewer (CG), and third reviewer (AK) independently reviewed the list of retrieved citations and abstracts in an un-blinded fashion and applied the selection criteria utilizing a standardized form (Appendix F). All retrieved abstracts were coded as per the following Agency for Healthcare Research and Quality (AHRQ) (2002) coding system (Table 3.2).
Table 3.2 AHRQ Coding Criteria

<table>
<thead>
<tr>
<th>Codes</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Include</td>
<td>Obtain full paper to assess</td>
</tr>
<tr>
<td>Ref-back</td>
<td>Obtain for reference or background information</td>
</tr>
<tr>
<td>Exclusions</td>
<td></td>
</tr>
<tr>
<td>ECL</td>
<td>Editorial, comment, or letter</td>
</tr>
<tr>
<td>NR-design/methods</td>
<td>Not relevant: design or methodological issues</td>
</tr>
<tr>
<td>NR-IA</td>
<td>Not relevant: implementation/application issues</td>
</tr>
<tr>
<td>NR-OCD</td>
<td>Not relevant: opinion/commentary/description</td>
</tr>
<tr>
<td>NR-ROS</td>
<td>Not relevant: report of study</td>
</tr>
<tr>
<td>NR-Review</td>
<td>Not relevant: review/overview</td>
</tr>
<tr>
<td>NR-Stat Meth</td>
<td>Not relevant: statistical methodology issues</td>
</tr>
<tr>
<td>NR- Other</td>
<td>Not relevant: other reasons/issues</td>
</tr>
<tr>
<td>NR- Text word only</td>
<td>Not relevant: studies identified had text word in title and/or abstract, but not relevant content.</td>
</tr>
</tbody>
</table>

(Agency for Healthcare Research and Quality, 2002)

After titles and abstracts that appeared to meet the inclusion criteria were identified, the full-text format of each study was retrieved for provisional inclusion. Articles identified by each reviewer were compared to ensure reproducibility of this stage.

The selection criteria were then re-applied to the full reports by the first reviewer (SB) and the second reviewer (CG). Decisions were made to determine appropriateness for a full critical appraisal; no disagreements between reviewers were encountered.

**Study Quality Assessment**

Incorporating the assessment of the quality of individual primary studies contained within a systematic review is an important step towards ensuring the validity of the review’s results. One recent study investigating the effect of incorporating poor quality research into meta-analyses determined that efficacy could be exaggerated by as much as 30-50% when results of lower-quality trials were pooled (Moher, et al., 2004). Quality assessment involves a critical appraisal of a study’s internal validity, in other words, the ability of a study’s design, conduct, and analysis to minimize bias or errors. Quality constructs vary, but typically include an assessment of: (a) study quality (methodological quality); (b) bias (systematic error); (c) internal validity (validity); and (d) external validity (generalizability/applicability). Study quality reveals the measures researchers have taken in their design, conduct, and analysis in order to limit bias. Bias is a systematic departure from the “truth”, while internal validity is the degree to which the outcomes of a study are likely to represent the “truth”. And finally, external validity is the extent to which
the effects observed could be applied in the real world based on the population, interventions and outcome measures utilized (NHS, 2001).

While most researchers would agree that using only methodologically sound research would limit bias, the exact process by which quality assessment is carried out varies. Numerous quality assessment instruments have been designed to help facilitate this step (AHRQ, 2002; Kmet, Lee & Cook, 2004; Law et al., 2005; NHS 2001, NHS 2002). One type of tool utilizes a scale format whereby points are awarded based on the presence of criteria deemed essential for study validity; the criteria included can either be empirically supported or arise from best practices in research design and conduct. In some cases each criterion is weighted differently depending upon its perceived importance. Since a final "quality score" is the end product of an assessment, quality scales provide a quantitative indicator of a study’s quality. The difficulty with quality assessment scales is that scoring can be somewhat subjective. Another type of quality assessment system utilizes a checklist format whereby quality items are listed but not scored numerically. This method provides a qualitative assessment of a study. These scales and checklists can include from 3 to 57 items and can take anywhere from 10 - 45 minutes to complete.

Quality assessment systems have been criticized for confusing the quality of reporting with the actual validity of the design and conduct of the trial itself (Alderson, et al., 2004). Hence, there is a chance that a well-designed trial that is poorly reported could be rated as having low quality; or, conversely, a poorly designed study that falsely or insufficiently reports all the required criteria could be rated as high quality. Ensuring comprehensive trial reporting is a well-recognized issue and one that is being rectified through standardized reporting guidelines. For example, in 1996 a group of clinical epidemiologists, biostatisticians, clinical trialists, and biomedical editors published a statement called CONSORT (Consolidation of the Standards of Reporting Trials) (Moher, et al., 2004). This statement contains a checklist and a flow diagram to assist authors in providing journal editors with the required information for publication of RCTs. The CONSORT statement was further updated in 2001 based on emerging evidence on the importance of various elements of RCTs (Moher, Schulz & Altman, 2001). The QUOROM statement (The Quality of Reporting of Meta-Analyses) followed shortly thereafter (Moher et al., 1999). While essentially designed for meta-analyses, QUOROM also has utility in the reporting of systematic reviews of RCTs. In addition, guidelines have subsequently been developed for reporting meta-analyses of observational studies (MOOSE) (Stroup et al., 2000); diagnostic studies (STARD) (Bossuyt, et al., 2003; Moher, Altman, Schulz & Elbourne, 2004); the review and monitoring of RCTs by research ethics board (ASSERT) (Mann, 2005); and the reporting of observational studies (STROBE) (Strobe Group, 2005).

It is for the reasons described above that the Cochrane Collaboration cautions reviewers planning on using such systems in their research. Their position is that since there does not exist a
"gold standard" upon which to compare quality assessment systems, they are not empirically based, and can be time-consuming and potentially misleading. They recommend a simpler approach that rates relevant criteria as "met", "unmet", or "unclear" and then summarizes these results in order to obtain an overall assessment of validity by classifying each study as "low risk of bias", "moderate risk of bias", or "high risk of bias".

On the contrary, other major research agencies support the use of quality assessment instruments. While the U.S. Agency for Healthcare Research and Quality (AHRQ, 2002) recognized their utility, they cautioned that the chosen tool must match the study design. They emphasized that quality assessment systems are not "one size fits all", and that they are unique to the specific study design they are being used to assess. While some instruments are touted as being efficient at assessing both RCTs and observational studies, the AHRQ does not recommend this approach as it may measure study quality less precisely.

The U.K. NHS Centre for Reviews and Dissemination (NHS, 2001) is another agency that endorses quality assessment instruments. Their position is that the individual quality components of checklists are superior to a final quality score obtained by a scale. It is important to note however, that tools are only an aid to critical appraisal and that their scores or values should not replace informed and thoughtful judgments (School for Health and Related Research, 1996).

**Quality Assessment Scale for RCTs**

The literature review uncovered two commonly used quality assessment instruments for RCTs: (a) Chalmers (Chalmers, Smith, Blackburn & Silverman, 1981); and (b) Jadad (Jadad et al., 1996). While both tools have demonstrated similar utility in quality assessment, the Jadad may be preferable as it is simpler and quicker to use (5-10 mins for the Jadad vs. 60 mins for the Chalmers) (Ohlsson & Lacey, 1995). The Jadad Quality Scale was originally developed in 1995 for the field of pain research. Its final version comprised three items: (a) randomization, (b) blinding of the patients and the investigator to the treatment, and (c) a description of losses to follow-up. The Jadad instrument is the only scale utilizing standardized scale development procedures that have been tested for inter-rater reliability. This scale assigns a score from 0 to 5 to each of the three criteria; the higher the score the better the quality of the clinical trial.

The Jadad scale was later modified to score an additional three items: (a) a description of inclusion/exclusion criteria, (b) a description of the method used to assess adverse effects, and, (c) a description of statistical analysis. This Modified Jadad was also evaluated for inter-rater reliability in a systematic review of Alzheimer's disease drug trials, obtaining a correlation coefficient of 0.90 (Oremus et al., 2001). The results of that study suggest the Modified Jadad's extended utility outside of its original application in pain research. In fact, the literature review found that the Jadad Quality Scale had been used in a wide variety of systematic reviews from acupuncture (White & Ernst, 1999) to Chinese herbal medicines (McCulloch, Broffman, Gao & Colford, 2002) and upper extremity disorders (Gummesson, Atroshi & Ekdahl, 2004).
scale overcomes some of the objections raised by the Cochrane Convention not only by scoring the presence of a specific criterion, but also by assigning an additional point if the criterion had been implemented appropriately.

The Modified version of the Jadad Scale (as opposed to the original version of the scale) was selected for this thesis based on the judgment that the additional criterion—"a description of inclusion/exclusion criteria"—was essential to the quality assessment process since one of the criticisms of Lovaas' research was the unrepresentativeness of treatment groups. However, two criteria of this modified version were not relevant to behavioural research. The first was whether the study was double-blinded, as obviously children and therapists cannot be blinded to treatment (however, outcome assessors can be). Therefore, the question was re-worded as "Was the outcome assessor blinded". And, the second criterion that was irrelevant to Lovaas studies pertained to adverse effect assessment; therefore this criterion was deleted. Therefore, the highest score obtainable was now seven (Appendix G).

**Quality Assessment Checklist for Observational Studies**

The Reisch et al. checklist (Reisch, Tyson & Mize, 1989) was used to appraise the quality of retrieved observational studies. Although this tool can also be applied to RCTs, based on recommendations that a tool should be specific to a particular research design, it was employed only for observational studies. The guidelines for implementing this tool were originally published in the journal *Pediatrics*; however, the article did not include the actual Checklist. Therefore, the author was contacted in March of 2005 in order to obtain a copy. This tool was the only system evaluated by the AHRQ that included all five of the critical domains for observational studies: (a) comparability of subjects, (b) exposure or intervention, (c) outcome measure, (d) statistical analysis, and (e) funding (AHRQ, 2002). The checklist included the following criteria: (a) purpose of study, (b) experimental design, (c) sample size determination, (d) description and suitability of subjects, (e) randomization, (f) stratification, (g) control use, (h) procedures for treatment/management, (i) blinding (masking), (j) subject attrition, (k) presentation of data, (l) analysis of data, (m) evaluation of subjects and treatment/management, and (n) recommendations/conclusions (Appendix H).

The authors of the Reisch et al. 1989 article note that the Check List was subsequently revised after obtaining feedback from medical residents and fellows. Table 3.3 outlines the criteria that were omitted from its original version.
Table 3.3 Reisch et al. (1989) Quality Assessment Tool: Revised Criteria

<table>
<thead>
<tr>
<th>Section(s) (1989 version)</th>
<th>Omitted criteria (2005 version)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Purpose of study</td>
<td>E: Source of support</td>
</tr>
<tr>
<td>3) Sample size determination</td>
<td>A1c: Use of expert monitoring committees for sequential design studies.</td>
</tr>
<tr>
<td>4) Description and suitability of subjects</td>
<td>A6: Predefined exclusion criteria D: Whether subjects are unsuitable for the questions being investigated.</td>
</tr>
<tr>
<td>5) Randomization and Stratification</td>
<td>(This section was divided into two sections in the 2005 version). Selection bias; evaluation bias (from un-blinded examiners); unrepresentative sample.</td>
</tr>
<tr>
<td>7) Procedures for treatment management</td>
<td>A: Informed consent</td>
</tr>
<tr>
<td>11) Presentation and analysis of data</td>
<td>(This section was also divided into two sections in the 2005 version)</td>
</tr>
</tbody>
</table>

Quality Assessment Process

Two researchers (SB & CG) independently appraised the retrieved studies in an un-blinded fashion utilizing the relevant quality assessment tool. The critical appraisal included relevant assessment scores and descriptive elements for each study critiqued. The “consensus agreement” was recorded in red ink on the respective quality assessment form.

Data extraction. Data extraction is the process whereby the reviewer retrieves information essential to a critical review directly from the study report. In order to limit bias in this somewhat subjective process, a data extraction tool with clear instructions was used (Appendix I). Results of the data extraction were discussed with the 2nd and 3rd reviewers.

Critical Appraisal

A critical appraisal of effectiveness evidence requires an appraisal of the quality of each study and then a conclusion on the strength of the evidence on the body of knowledge as a whole. Critical appraisal techniques evaluate not only internal validity, but also external validity as well. A thorough appraisal can explore quality differences that may explain the heterogeneity of study results, which can aid in the interpretation of outcomes to be used to inform practice and research. The three objectives of critical appraisal are: (a) to understand the rigor of the studies, (b) to uncover reasons for differences among study results, and (c) to
provide readers with the required information to judge the applicability of the review to their clinical practice (Meade & Richardson, 1997).

After examining previous systematic reviews of Lovaas therapy (Bassett et al., 2000; ECRI, 2000) it became evident that it would not be feasible to perform a meta-analysis for this thesis. This was based on the identified clinical diversity (clinical heterogeneity) inherent with Lovaas Autism Treatment, the participants, and the clinical outcomes. This decision was supported by guidelines developed by the Canadian Agency for Drugs and Technology in Health (CADTH), formerly the Canadian Coordinating Office for Health Technology Assessment (CCOHTA). These guidelines stipulate that the finding of heterogeneity in retrieved studies would steer the systematic review towards a qualitative assessment without pooling (CCOHTA, 2003).

Data extracted from the retrieved studies were individually appraised for validity utilizing criteria developed by the National Institutes of Health (Bristol et al., 1996). Utilization of these criteria provided continuity with the critical appraisal undertaken by the researchers in the original BCOHTA report of Lovaas effectiveness (Bassett et al., 2000).

After the retrieved primary studies and systematic reviews were critically reviewed, descriptive data synthesis was then employed to collate and summarize the studies' characteristics and results. NHS (2001) identifies the key elements of descriptive data synthesis as including the following characteristics:

1. Population.
2. Interventions.
3. Settings where the technology was applied.
4. Environmental, social and cultural factors that may influence compliance.
5. Nature of the outcome measures used, their relative importance and robustness.
6. The validity of the evidence.
7. The sample sizes and the results of the studies included in the review.

Along with describing the study, descriptive data synthesis enables researchers to assess whether the characteristics of the participants, interventions and outcomes of the studies would permit the eventual generalization of their systematic review results. This method also allows an assessment of the quality of the studies in order to determine the trustworthiness of the results. In addition, when comparisons are planned, descriptive data synthesis has the ability to identify any missing data or clinical heterogeneity that would prevent use of the data for meta-analysis purposes. However, even if a quantitative analysis cannot be conducted, descriptive data synthesis facilitates assessment of whether the treatment effect is large enough to be regarded as obvious, and if so, whether these effects are consistent across the analyzed studies (NHS, 2001).

The extracted data were summarized in tabular form in order to qualitatively assess the studies for significant differences in the key characteristics of participants, interventions, and
outcome measures (clinical heterogeneity); the study designs and quality (methodological heterogeneity); and the reported effects (heterogeneity in results) (NHS, 2001). In addition, limitations were identified that resulted from missing information from the included studies.

**Summary of Research Design**

This chapter has detailed the three methodologies utilized within this thesis – Grounded Theory, Systematic Review, and Critical Appraisal – along with their methods of application. As previously mentioned, the objectives of this research are:

1. To examine the role of evidence within the contexts of law, health policy and health care.
2. To determine the current state of knowledge in regard to the effectiveness claims of Lovaas therapy for autism spectrum disorder.

In order to achieve these goals, it was necessary to design a mixed methodological approach that incorporated multiple methods (Figure 3.2).

**Figure 3.2 Study Design: The Legal and Effectiveness Contexts of the Comprehensive HTA Framework**

<table>
<thead>
<tr>
<th>Legal context</th>
<th>Effectiveness context</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data source:</strong> Participants &amp; documents</td>
<td><strong>Data source:</strong> Primary/secondary research</td>
</tr>
<tr>
<td><strong>Data collection:</strong> Interviews and document analysis</td>
<td><strong>Data collection:</strong> Systematic search</td>
</tr>
<tr>
<td><strong>Data analysis:</strong> Grounded theory</td>
<td><strong>Data analysis:</strong> Systematic review &amp; critical appraisal</td>
</tr>
<tr>
<td>- Evidentiary Process (Figure 2.1)</td>
<td></td>
</tr>
<tr>
<td>- Scientific Evidence Pathways (Appendix B)</td>
<td></td>
</tr>
</tbody>
</table>

The methods utilized within the effectiveness context of the HTA framework have been well established; therefore guidelines for the design, conduct and analysis were closely followed. However, the HTA literature revealed a paucity of methods for studying (and understanding) the legal context. Therefore, a review of qualitative options was undertaken. Content Analysis was the first approach reviewed for its fit to the research question. Content Analysis is, "a quantitatively oriented technique by which standardized measurements are applied to metrically defined units and these are used to characterize and compare documents" (Manning & Cullum-Swan, 1994, p. 463). This research approach establishes the existence and frequency of concepts found in text and can examine relationships between concepts. The problem with Content Analysis (and the reason this methodology was not chosen) is that the
researcher must have an idea of the types of concepts to look for *a priori*. Although grounded theorists can begin their research with general concepts abstracted from the literature, the methodology still provides opportunities for concepts to arise from the data itself. Content analysis has also been criticized for its lack of sensitivity to the context of the text. Grounded Theory was the second qualitative approach investigated. This methodology was considered appropriate since little is known about this topic: "[Grounded theory] supports the theorizing of 'new' substantive areas which can be further explored to develop a more formal theory" (Hunter et al., 2005, p. 66). In addition, there is a good conceptual fit as the origins of grounded theory are rooted in my professional background of nursing.

Collectively, these three methodologies were systematically applied to the data in order to determine how the various domains conceptualize evidence and also to discern the current state of knowledge in regard to the effectiveness evidence on Lovaas Autism Treatment.
CHAPTER 4: RESULTS

Analysis of the Role of Evidence

The first objective of this thesis was an examination of the differing conceptualizations of evidence by legal, health policy and health care sectors. This analysis was structured around my conceptual framework (Appendix B & Figure 2.1), developed from work done prior to the Auton case (Eisenberg, 2001), and embedded within the context of judicial policy making (Auton and Anderson legal proceedings). Data were obtained from interviews and legal documents and was analyzed using grounded theory methods (Figure 3.2).

The descriptive analysis of the seven original conceptual categories (Appendix B) resulted in a refinement of my original conceptual framework. The analysis revealed that the contextual category Level of Evidence related better as a subcategory of the evidentiary actor category decision-makers. In its new position in the hierarchical tree, it defines one of the properties of decision-making, which is determining the level of evidence to apply to an issue. In addition, the evidentiary actor category of change agent overlapped significantly with decision-makers; therefore, these two were merged with the label remaining as the latter. In addition, gatekeeper was relabelled as adjudicators. Change agents ended up describing more the process of change as opposed to who initiated the change; therefore, it was redefined as a contextual category and termed process of change. And finally, level of application of evidence was changed to focus of application of evidence to more aptly describe its properties. Figure 4.1 is my revised depiction of the evidentiary process.

Figure 4.1 The Context, Actors and Focus of Application of the Evidentiary Process (Revised)
These next sections will provide a description of the key concepts I discovered through the coding and analysis of the data (summarized in Appendix J), and will conclude with a re-conceptualization of my scientific evidence pathways framework (Appendix L).

**Descriptive Analysis of Conceptual Categories**

Analysis began with a sentence-by-sentence examination of the data in order to identify the segments of text that were empirical indicators of the categories in the framework, along with some of their unique properties and dimensions. In addition, the text was also reviewed for references to the broader concepts of seeking, understanding and applying evidence. After the first round of open coding, 44 free nodes (open categories) and 64 tree nodes (hierarchical categories) were identified. Axial coding later consolidated and refined the themes into six main categories: (a) perspective of evidence, (b) rules of evidence, (c) process of change, (d) adjudicators of evidence, (e) decision-makers of evidence, and (f) focus of application of evidence. These six categories, applied to the three contexts of law, health policy and health care, resulted in a matrix of analyses consisting of 18 cells. The findings of these analyses are presented in the pages that follow.

**Perspective of Evidence**

The perspective of evidence is the first contextual variable and is defined as the type of evidence utilized for decision-making (e.g., population-based findings vs. individual-based, context-specific evidence) (Eisenberg, 2001). In addition, this category depicts the way each sector conceptualizes the patient/client – in either a population or individualist perspective. This concept was investigated through the posing of interview questions such as: (Appendix K)

1. “Does the application of evidence ever have an outcome for the population as opposed to strictly for the individual?”
2. “What about strong evidence in support of an intervention that would benefit a few people, vs. weaker evidence on an intervention that would benefit many?”

The text provides numerous examples of the legal domain's predominantly individualist perspective: “Each case is decided on the merits of what is presented and on the evidence that’s heard. So most cases are, of necessity, fact-driven in terms of what’s placed before the judge” (Participant: L1).

And also:

Most plaintiffs' lawyers wouldn’t worry terribly about what’s happening from a policy perspective, or what happens to other people and whether they get funding, what the funding envelope for their clients does for other clients - that’s not their concern. (Participant: L2)
And, specific to the Auton case, this quote speaks to utilizing evidence at an individual level, "I think that the issue of effectiveness [of Lovaas therapy] is a pretty concrete one, which is whether or not the children [italics added] seemed to benefit concretely from the therapy" (Participant: L2).

Alternatively, another legal professional stated the following:

It would be my view that the findings of the trial judge that Lovaas therapy is of significant assistance to some autistic children represented a public [italics added] and reasoned endorsement of Lovaas therapy in British Columbia, to a degree that it had not had before. And so, I think that actually is a significant public fact [italics added], which helped in the policy debate. I think the factual success that they had at trial in establishing by the judge that the therapy had credibility, ended up being a persuasive element in the public policy push to have increased funding made available. (Participant: L2)

A health policy professional cites a Pharmacare example of either providing an inexpensive treatment to a large population or funding an expensive treatment to a small cohort (e.g., an "orphan drug"): Those are very difficult decisions for governments to make, for policy-makers to make. And, if anyone figures out the magical solution as to how you do that in a totally objective and dispassionate way I don't know what it is because a lot of it comes down to trying to do the right thing. (Participant: HP1)

And, in regard to health care practitioners, the text proffers examples of utilizing population level research but customizing it to an individual patient:

You know, there's no good clinician [who] will just randomly apply just, you know, like a sort of bottom line from their research article. You have to meld the two; you have to have a good understanding of what that group of data means; the limitations of that; and then have a really good understanding of your client or your patient. (Participant: HC2)

**Rules of Evidence**

Rules of evidence is the second contextual category. This category is defined as an agreed-upon set of rules used to accept or reject evidence. Interviewees were asked questions such as:

1. "What rules of evidence do you follow?"
2. "How do you decide what evidence is acceptable?"

The legal data were full of references to formalized rules of evidence such as the rules for admissibility:
The issue of course is first of all, whether it's a matter that is outside of the normal knowledge and understanding of the judge, and is it something that a judge would be assisted about [sic] by receiving the opinion of someone who, by reason of study, discipline, or experience has a more reliable opinion than a lay person or a judge in that case (Participant: L1)

Since, under the adversarial system, the litigants have the right to their “day in Court”, and judges are assumed to be unfamiliar with the substantive backgrounds of experts, leniency is granted to the parties and their lawyers in regard to presenting evidence and calling witnesses. In the end, though, the weight of the evidence is still determined by the trier of fact (jury or judge sitting without a jury). Pertaining to the choice of evidence to lead at trial, the data appeared to indicate that that process could be somewhat selective:

The defence counsel gets to wait and see what the other sides have got [sic]; you don’t have to tell the other side everything . . . just the parts that are going to help your case you disclose, in terms of expert evidence. (Participant: L1)

The legal data showed that legal rules of evidence evolve over time:

It used to be that the standard of care would be determined by what the average person in a similar calling with similar experience would do. More recently the Courts have come to accept that as long as there's a recognized school of thought within a profession that supports the conduct of the individual, you don’t have to be average and it simply has to meet the approval of an accepted school of thought. So, arguably you don’t have to get to the 50% anymore. (Participant: L1)

Within the health policy realm, there was a generalized awareness of a structured review process for pharmaceutical evaluation. One health policy professional identified a condition that impacts the acceptance of evidence into the political forum:

There’s an art in the science of being a good policy person . . . you have to be able to read the culture that you are in . . . in some governments . . . carefully collected, scientifically rigorous public opinion would be viewed as unimportant - because the public, you know, didn’t have anything useful to say. So in that case, I would never put forward that kind of information as evidence for any particular opinion or position. So in certain cultures, political cultures, you come to see, to learn, what evidence is paid attention to and put that forward. (Participant: HP2)

Within health care, the practitioners did not articulate any formal rules of evidence. Instead, they indicated that research interpretation is based on their training in appraisal techniques. These skills are maintained through processes such as continuing to read up on research, and participation in formal review committees. No accepted appraisal tool is used; instead, one practitioner (Participant: HC1) refers to a mental “template”. In general, these clinicians attested to possessing a basic knowledge of research appraisal techniques.
Process of Change

Process of change is the third contextual variable and is defined as the differing ways in which change occurs and how it impacts each respective sector’s conceptualization of evidence. This variable was investigated by posing the following types of questions:

1. “How do changes in your clinical/professional practice come into effect?”
2. “Who are the ‘change agents’ of your profession?”

This category proved rich in textual references. The ensuing analysis uncovered the following subcategories: (a) level of evidence required to implement change at the practitioner level, (b) strategies for change, (c) reasons for change, (d) barriers to change, and (e) facilitators of change.

**Level of evidence required to implement change at the practitioner level.** Only the health care sector made reference to this category. One practitioner (Participant: HC2) felt that “several articles on the topic from the right centres and the right methodology and good statistics” would be required before a change in practice would be contemplated. They discounted expert evidence (e.g., a “guru” in the field) as prompting change; instead, they hypothesized that experts might serve to change thinking and ideas towards their patient population, but rarely towards their actual practice.

**Strategies for change.** All three sectors spoke of strategies for change. For the legal domain, the interview questions prompted explanations of systems to appeal outcomes of trials and the resultant changes to common law.

The reason lawyers continue to go to Court is because interpreting the judgments from the Supreme Court of Canada, or in the appellate Courts, isn’t necessarily a simple task in that there are events and situations that arise that nobody anticipated. (Participant: L1)

A health policy professional stated that, “getting an issue onto the public agenda and the political agenda [was] really important [as] to where the policy directions go” (Participant: HC2). Health care’s strategy for change includes getting involved, consulting experts, involvement in private practice, reading literature, participating in research, and embarking on professional development workshops. Another health care practitioner’s thought on how to stimulate change was to “shake things up by showing them something different that maybe people haven’t been aware of or haven’t been thinking [about] outside the box” (Participant: HC1).

**Reasons for change.** Only the domains of health policy and health care proffered opinions in regard to reasons for change. Within the health policy context, one professional sensed that change occurs in order to keep pace with new political parties, “as political parties change or political priorities change, different kinds of policy options are considered and different kinds of policy directions are pursued” (Participant: HP2). Within health care, requirements of educational institutions to maintain academic appointments, along with...
licensing requirements from professional bodies were seen as the main reasons for change. In addition, clinical experiences where patients do not match textbook presentations, or moments of “inspiration” to adopt new practices, were also viewed as potential reasons to change. A couple of interesting in vivo codes (verbatim codes) that came up were, firstly, a reference in the health care data to practitioners being “influenced” to change; and secondly, a description in the legal data to lawyers “persuading” change.

**Barriers to change.** The textual analysis identified specific barriers to change within the sectors of health policy and health care. Not surprisingly, economic barriers to change were recognized by the health policy actors:

> When you wind up with conflicting evidence or even conflicting opinions as to whether the evidence that shows that a dollar will be saved down the road, those [policy changes] are much more difficult to implement (Participant: HP1).

Along with, “As the economic cycles go up and down . . . when you’re in a period of restraint, it doesn’t matter how good your evidence is” (Participant: HP2).

Within the health care context, practitioners who work within private practice were seen as being at a disadvantage due to long hours with limited exposure to outside influences of change. In addition, the practitioners were of the opinion that typically, people do not want to change.

Another finding from the health care data was reluctance on the part of the practitioner to discredit alternative treatment therapies:

> We hate to be telling parents, you know, you’re nuts and there’s nothing to that [treatment]. And so, there might be something to it, so you know, we’re always a little cautious to be the one saying, “You’re all crazy” and we want to wait for the evidence to come in; and that seems to take a long time. (Participant: HC2)

In fact, this excerpt could be defined as either a barrier to change (as it has been here) or as a facilitator of change (as it will be in the next section), depending upon the context of the action. For example, if a parent discloses to the psychologist that they have placed their child on a gluten-free diet, by not discrediting this treatment (and perhaps recommending something else) change does not occur (e.g., the patient stays on the diet). Alternatively, if the parent first solicits the psychologist’s advice on the topic, then, by not discrediting it, change may occur (e.g., the parent may put their child on the diet).

**Facilitators of change.** Within the legal forum, the data analysis picked up that some litigation victories are quickly forgotten and have no impact on society. Alternatively, “litigation failure [could] result in a change of approach because of some of the facts and findings that come out of the case” (Participant: L1).

In regard to the health policy sector, political factors and stakeholder groups were identified as facilitating change, “in my experience it is the public agenda that drives a lot of
policy at a macro policy level” (Participant: HP2) (Evidence was identified as the facilitator of change at the micro policy level). And finally, both economic factors and litigation remedies were seen to shape policy.

Within the health care field, although private practice had been previously identified as a barrier to change, other types of practice settings such as educational institutions and hospitals were recognized as facilitators of change. Factors such as access to a variety of experiences, research and professional development opportunities, along with additional time and money, were amongst the key facilitators. Personal motivation, access to experts, impact of the media, in addition to patients themselves, also emerged as important variables. And, as previously described, the finding that some health care professionals were reluctant to discredit alternative treatments was also a facilitator of change. One final interesting passage from the text speaks to patterns of utilization of evidence, "I really would only use evidence probably when it comes to making recommendations. I don’t think they’re a good influence for what I actually [do] with the child" (Participant: HC2).

**Adjudicators of Evidence**

The first “evidentiary actor” category – adjudicators – required an examination of the text for indicators of how each sector used experts to adjudicate differences (in regard to evidence). This category was operationalized by posing the following types of questions:

1. “How do you assess expert witnesses’ qualifications?”
2. “How do you evaluate the reliability and weight of expert evidence?”
3. “Does the peer review process, and journal editors in particular, act as the gatekeeper of evidence for your profession?”

The textual analysis identified five subcategories between the sectors of law and health care: (a) experts used to provide opinion, (b) experts used to discredit opinion, (c) experts used to conduct research, (d) experts used for peer review, and (e) Court-appointed experts.

**Experts used to provide opinion.** The most frequent reference in the legal data was in regard to expert witnesses. The main purpose of this role is to present an opinion within their area of expertise to the Court on a matter that is outside the common knowledge and experience of the trier of fact, and to provide their opinion on the facts of the case so as to inform the Court on an issue beyond its understanding, e.g., what is medically appropriate and what is not. Specific to the search for an expert witness, one legal professional proffered their approach:

There are certain usual suspects [experts] that we know in certain areas, and if it's an area you’re familiar with, you know who the "go to" witnesses are and you know who helps one side more than the other; and, if you’re on that side you go get them and if you’re on the other side then you get a different list. (Participant: L1)

The ideal expert witness was identified as the primary author of a study; however, if the report was considered seminal in its field, secondary sources of opinion were also acceptable. The
You have to ask them [expert witnesses], first of all, if they recognize the literature and if it's considered reputable and reliable in their area. If they say "no", that's the end of it. You don't get to rely on that particular piece of literature unless somebody else is prepared to dignify it. (Participant: L1)

In addition to:

[T]o some extent, you're flying a little blind here that you are relying on somebody else to tell you whether a study is good, bad or otherwise; and they might be right, but if they can't articulate it in a way that's persuasive, the judge may not agree, and so it's what I said earlier - it's important not just to get someone with good credentials, you got [sic] to get somebody who can communicate effectively. (Participant: L1)

Legal participant L1 also suggested that, some lawyers prefer to pre-screen expert witnesses prior to trial:

Well that's part of the art of advocacy is to try and make somebody look biased . . . and maybe they are and maybe they aren't. I don't mean to be facetious. I mean obviously lawyers aren't supposed to be misleading Courts, but when you have an expert, as I've said, you're entitled to know who they are and what they've done and what they're going to talk about before they get there. And there are a variety of search engines that exist now that allow you to go and look and see if they've given evidence in other cases. And if they have, what they've said and whether the judge has liked what they had to say or not. Um, certainly anybody who gives evidence in British Columbia we can check on Quick Law. There are similar search engines in other jurisdictions . . . and as well, if it's an area that is a controversial one, you can find lawyers in other jurisdictions who have some expertise in those areas, [and] they'll happily give you transcripts of evidence and often reports that some of these people have written. (Participant: L1)

There were limited textual references to experts used to provide opinion in both the health policy and health care data. Health policy professionals acknowledged the commissioning of researchers to provide opinion, while the health care sector utilizes peer reviewers (to be discussed in a subsequent section).

Experts used to discredit opinion. The text also revealed various processes utilized by lawyers during cross-examination to contradict or discredit evidence:

So we know ahead of time the qualifications of the experts that I'm going to have to deal with. And often, for the opposing experts, I'll look at their list of publications and if I think that there's an article that I might be able to use to persuade them that they've made a mistake, I'll look at that. (Participant: L1)
In addition to:

If you’re attacking their opinions, you either attack the factual matrix upon which their opinion is based because if they get the facts wrong, the opinion probably won’t get any weight. Or, you attack the opinion itself; but, that’s a far more difficult proposition for a lawyer for sure – playing in somebody else’s ballpark. (Participant: L1)

A good example of this approach for discrediting an expert was found in the text of Auton (2000):

Dr. Gresham expressed the view that because intensive discrete trial training is not provided by doctors, it is therefore not a medically necessary treatment for autism. However, he readily conceded that, as he is not a physician, he is not qualified to offer that opinion. (para. 33)

However, despite lawyers’ arsenal of interrogatory tools, there were also indicators of their limitations:

In terms of medical evidence, I think it’s a mistake usually for the lawyers to think they’re going to persuade the judge as to what medically is appropriate or inappropriate. That’s for the expert witnesses. (Participant: L1)

There were no data for this category from the health policy domain, while the health care data once again referred to peer reviewers.

**Experts used to conduct research.** Another method of utilizing experts in the process of evidential adjudication is to commission them to conduct additional research. This example was found in Auton (2000) and pertained to the BCOHTA report. Within law, since the research has already been conducted the role of the expert is to report on it. In health policy, if there is sufficient lead-time, research studies can be initiated, however, this is an expensive prospect and infrequent occurrence. However, in the health sector, research is an inherent activity of the profession.

**Experts used for peer review.** References in the health care data described some of the benefits of peer review including, having an expert examine and critique a manuscript (albeit, sometimes not very thoroughly), and receiving constructive feedback and assistance with research design issues. Adhering to reviewers’ feedback was identified as being essential in order to get a paper accepted; there was also a reference to specific types of research that typically get published:

I’m sure the system can be improved in that there are things that are published because they’re sexy or because they’re, you know, as I said, the editor thinks it’s maybe important and the editor happens to be a leader in the field, and, null findings are not very sexy, nor are replication studies not very interesting to most people. (Participant: HC1)
In addition:

I don’t know if the peer review process is the very best, but I think it’s reasonably fair and having two or three people review something, I mean, it’s probably not as extensive as it should be but I think nothing would ever get published if you had about ten people weighing in on a particular article. But as gatekeepers, I think it’s probably a good thing. (Participant: HC2)

An interesting dimension of peer review that was also identified in the health data was in regard to a journal’s reputation:

But, you know, journals have reputations too, not just people. And so I guess the ones to have had [sic] fair gatekeeping, and maybe scientists are kind of cautious too, and they don’t want to deviate too far. And so, if a particular journal is known for their editorial policy of being a little bit more cautious, I think maybe we’d put more faith in it, even if they then may not come up with a great cure or they’ll be slower to. (Participant: HC1)

There was no text reference to peer review in either the legal or health policy data.

**Court-appointed experts.** Experts can be appointed by the Courts to aid judges with the interpretation of evidence during trials. One legal professional noted, “There has been an assessor sit with the judge; we do have that ability, but that person wouldn’t give evidence - they would listen to evidence and help determine for the judge” (Participant: L2). However, another professional did not view this process favourably:

There has been some suggestion that the Court be able to force joint experts on parties. I think it’s inconsistent with the adversarial process to ask somebody to share when their interests are opposing. But, there is a view that that would be a step in a positive direction; this is not a view that I happen to share. (Participant: L1)

In addition,

It’s my experience that a Court appointed expert is often simply a way of purchasing bias without recognizing it as bias, because you have someone who has the immediate access to a judge and their view may be disproportionately influential, and they may come from one or the other school that’s represented in the opposing ranks of experts in the trial. . . and there’s a terrific risk that the trier of fact - the jury or the judge - come to regard their expert as somehow free of all the biases that the rest of us have, and you’ll be free of bias when you don’t talk to more than one person. (Participant: L2)

This category does not pertain to health policy or health care.

**Decision-Makers of Evidence**

Decision-makers are the other “evidentiary actors” in this framework. The data analysis explored who determines which evidence is to be followed. Once again, this was pursued by the following questions:
1. "How do you decide what evidence is admissible?"
2. "Who decides on the validity of the evidence?"
3. "How do you decide what evidence is acceptable?"

The data analysis revealed five subcategories for decision-maker: (a) key players, (b) types of evidence, (c) level of evidence, (d) barriers to decision making, and (e) evidentiary basis for policy making.

**Key players.** Not surprisingly, the legal data validated the role of judges and lawyers as decision-makers of evidence. "At the end of the day, it's the judge who decides [admissibility of evidence], how much weight any particular piece of evidence gets, how credible one witness is as opposed to another" (Participant: L1). Their role includes determining the validity of the evidence and then basing their decision strictly on the evidence led by the lawyers. One text reference pointed out that sometimes judges overstep their boundaries:

"It's not their job to start wading into the evidence and cross-examining witnesses. Certainly my view is that they ought not to be doing it, and when they do, it can be difficult because there are many lawyers who don't feel they can confront a judge by saying, "You're not allowed to do that"." (Participant: L1)

Judges interpret existing law and its application, including the rarely exercised power to reject professional standards (e.g., standards of care) if determined to be unsafe.

Not surprisingly, retrieved passages of text from the various Reasons for Judgment were characteristically objective and fact-driven. However, occasionally judges did interject their opinions. An interesting excerpt, which originated from Auton at the Supreme Court of Canada (2004), appeared to be offering advice to legal colleagues:

"Had the situation [Auton] been different, the petitioners might have attempted to frame their legal action as a claim to the benefit of equal application of the law by the Medical Services Commission . . . " (para. 45)

And, another selection from the trial judge in Auton (2001) presented the following interpretation of the issues before the Court:

"I venture the opinion, albeit with some hesitation, that the Government's failure to at least include Lovaas Autism Treatment in its arsenal of effective treatment techniques appear to result from some antipathy to the petitioners and the vehemence and effectiveness of their cause. (para. 25)"

While judges have the ultimate authority over the weight given to the evidence presented at trial by judge alone, initially it is the lawyers responsibility to decide what evidence to present to the judge; after that, their purpose is to convince the judge of its relevance, "My job is to try and persuade the judge of a trial fact, and [that] they should accept the evidence from the focus that I invited to the party" (Participant: L1).
Also in relation to key decision-makers, the text described the legal hierarchy of the Canadian Court system with a decision from the Supreme Court of Canada overruling contrary lower Court decisions.

Although juries were also identified as decision-makers of evidence, one legal professional (Participant: L2), stated that most medical cases are too technical for juries; hence the cases are typically tried by judge alone.

The key players in the health policy field were identified as the politicians and the public stakeholders. The text made reference to the fact that, as an elected official, politicians are the final decision-makers. Although evidence is utilized in decision-making, it was pointed out that it is only a small portion of the equation:

> When the politician wants something, ultimately our elected officials make the decision and so the level of decisions that they make, you know, they make those decisions and evidence is only one part of it; is only one factor that they consider (Participant: HP2).

In response to my question as to whether a politician could go against the evidence, the health policy professional responded, “Yes, they could and they might, and that would happen. So that’s what living in a democracy is all about” (Participant: HP2). Another health policy professional (Participant: HP1) acknowledged that frequent portfolio changes increased reliance on others and reduced the ability to ask critical questions about issues.

And, in regard to health care professionals, opinions were divided in regard to the utility of journal editors assuming the role of decision-makers and gatekeepers of the evidence for their profession:

> Some editors tend to publish what they think or like more than looking at the objective [of the research]. I mean, I have always kind of wondered why we don’t have more studies that have no findings, for example. Or, why we don’t have replication studies, you know, as much as we probably should. (Participant: HC1)

And, “There’s probably some truth to that [editors acting as gatekeepers], but it doesn’t bother me. I mean I’m glad. But as gatekeepers, I think it’s probably a good thing” (Participant: HC2).

Some journals were also viewed as having reputations for “fair gatekeeping” and being cautious with their publications. The thought was that cautious journals might be slower to publish studies with impressive outcomes and that perhaps cautious journals encouraged scientists to be cautious as well.

**Types of evidence.** Interspersed throughout the data were references to specific types of evidence utilized by the respective sectors. While not an exhaustive list, the interviewees make reference to the following:

- **Law:** affidavits, depositions, anecdotal, hearsay, standard of care, scientific, legislation, and precedent cases.
- **Health policy:** effectiveness, efficacy, and cost effectiveness evidence.
Health care: the full hierarchy of evidence

**Level of evidence.** This subcategory relates to whether the sectors utilize evidence from a single primary research study (e.g., an RCT) or from secondary synthesized designs (e.g., systematic reviews). Although no specific interview questions were designed to discern this information, the data did yield some information on this variable.

The legal domain made reference to preferring primary evidence to secondary evidence, but acknowledged the utilization of both forms.

Health policy professionals identified government databases (e.g., Pharmanet & Medical Services Plan) as an under-utilized source of evidence and suggested that this is due to "institutional resistance" (which is slowly being overcome). Another comment recommended determining the underlying purpose for submitted evidence, "I think that we have to look at the source of evidence . . . you have to look at who's providing the evidence and what their motivations are" (Participant: HP1). Although policy-makers acknowledged the utility of evidence-based medicine, one individual also proffered an opinion that, "I would also think that we've got our own ability to verify that data from our own data sources if we just used them" (Participant: HP1). Another source of evidence was identified in relation to Pharmacare's listing of one particular drug over another:

There is a great pushback, which is industry driven [italics added], to try to come up with a collection of anecdotal evidence of people who were intolerant to the new listing.

And, I found that that is one example, I think, of evidence coming forward through some advocacy groups that really is not doing anybody a service (HP1).

While heath policy professionals also acknowledged the "hierarchy of evidence", the text revealed that other forms of evidence, such as public opinion, were also valuable:

So, you could have the most strong [sic] evidence in the world on an issue [and] unless you have public support for recognizing that issue is important, . . . and political willingness to do something about it, it doesn’t matter how strong the evidence is (Participant: HP2).

Evidence also comes from public input. In general, policy-makers identified their main source of evidence originating from secondary research.

The health care data did not uncover an allegiance to any specific form of evidence, nor were weights attached to any one design, "I think I use all of those [research designs], I don’t rely primarily on one. I don’t necessarily put more weight on one than another. It depends on the situation" (Participant: HC1).

**Barriers to decision making.** A review of the data identified some indicators of factors that impact the decision making process. Within the legal domain, Court decisions may be difficult to reach, as judges are reliant upon the quality of evidence advanced by the lawyers, "By and large, the judges are left at the whim of the lawyers in terms of what goes before them."

138
(Participant: L1). Also, Auton (2001) provided an example to support Manfredi and Maioni’s (2002) assertion that, “rights discourse narrows the range of policy alternatives” (p. 215):

The Crown objects to a direction that it fund treatment whenever, and to the extent that a medical practitioner or psychologist recommends it on the basis that such a direction would narrow legitimate policy options with respect to diagnosis and assessment (para. 4)

Within health policy, politicians were seen as being bombarded with evidence presented by multiple stakeholder groups, each lobbying for their own unique cause:

The challenge is to reach past those stakeholder groups and actually connect with patients and connect with the general public to hear some of their real concerns, because if you were to be purely reactive in terms of the voices that are coming at you, which are the stakeholder community, you would not hear what I think is some of the public’s real anxiety . . . that [reaching out to the public] does tend to get past some of the vested interest (Participant: HP1).

In addition, another barrier identified by policy actors was the institutional resistance to using Pharmanet and MSP data due to privacy issues.

There were no text references to this subcategory from the health care sector.

Evidentiary basis for decision making. This variable was identified in the health policy data from the following quote:

I don’t think that policy is made entirely on the basis of scientific evidence, although . . . there has been much more attention paid to it. But for some things, some major public policy shifts, we don’t have evidence and we never will probably, and we make big policy shifts. (Participant: HP2)

Focus of Application of Evidence

The final category of the framework pertains to the outcome of the evidentiary process, which is the focus of the application of the evidence - the level at which evidence is applied, either at the level of the population or the individual. Although primarily working within an individualistic perspective, there is an excerpt from the legal data that described a context in which a legal outcome would be applied to a population: “If a class action had been certified by the judge, then the judicial findings would be binding on the Province in respect to all members of the class” (Participant: L1). However, typically evidence that is tried at one level of the legal system would not be generalizable to the next level of the Court:

There are cases that reach the higher Courts – particularly the Supreme Court of Canada – where there may be a statement about the existing state of affairs in one particular area. But, I don’t think there are too many cases where the evidence is going to necessarily bind a different judge in a different case who doesn’t get that evidence or gets different evidence. (Participant: L1)
And yet, in regard to the "symbolic award" in Auton (2001), although not granted to all autistic children in BC, Allan J. acknowledged the population perspective of the original class action attempt by recognizing that "[the petitioners] have shouldered the financial and emotional burdens of litigation on behalf of, and for the benefit of, a wider community of autistic families" (para. 64).

The focus of health policy is inherently population based, so no specific text references were required to support this fact.

An example from the health care data showed that, although evidence is applied at the level of the patient, this sector has to extrapolate population level data to their patient/client:

Because often you're dealing with group data and so this group information doesn't tell you specifically about this individual person that you're working with, and so sometimes you take the general ideas and then you take a look at what you know about your individual client-patient... it's kind of a balancing act where you're saying, "Well, ok, how much of this general sort of group data applies to this individual?", and then, how much do I know about this individual that would lead me to tailor that particular intervention or technique to suit their needs? (Participant: HC1)

Judicial Policy Making

For this thesis, judicial policy making is the umbrella context in which all of the previously described concepts reside and interact (Figure 4.1). Text from Auton (2001) highlighted the trial judge's understanding of the impact of the Court's decision on policy: "This case raises significant public policy issues as to the respective roles of the judiciary and the legislature" (para. 26). In addition to: "one recognized purpose of constitutional remedies is to regulate and modify governmental behaviour" (para. 16). The trial judge continued:

The issues raised by the petitioners underscore the difficulties inherent in a process where the Court's finding of unconstitutionality is designed to change governmental behaviour. The effective treatment of children must be delivered within a framework that is necessarily constrained by the resources available and the need to allocate those resources equitably in response to competing demands. (para. 26)

Not surprisingly, interview questions that addressed these issues were rewarded with copious free-flowing responses from both the legal and health policy professionals (indicative of an obvious contentious matter). One legal professional had voiced their opinion on the appropriate division of responsibility between the Courts and the legislature:

I think ideally the legislature should be dealing with many of the changes that confront Courts. On the other hand, when it involves the interpretation of the existing law, then the Courts are best equipped to say well, the way that one's written, here's the result and the legislature can change the wording. (Participant: L2)
While another expressed concern as follows:

I think that the Courts need to decide what their role is with regard to shaping health policy going forward. And, I believe that, I think that, the Courts today have taken on too much. I think they've crossed the line. (Participant: L1)

And finally:

If the Courts decide that they are going to continue to encroach upon the legislature's responsibility when it comes to the development of health policy, then they need to become much more robust in their ability to evaluate evidence then they are today. And, it is not in a Court context of lawyers presenting evidence and calling expert witnesses, because that does not work in my view when it comes to setting health policy. (Participant: HP1)

**Scientific Evidence Pathways**

The results from the textual analysis culminated in a revision of the original conceptualization of the Scientific Evidence Pathway (Appendix B). This is presented in its newest form in Appendix L.

This past section – Analysis of the Role of Evidence – has examined this phenomenon through the lens of grounded theory within the legal dimension of the health technology assessment framework. A discussion of the significance of these findings in the context of how the three sectors of law, health policy, and health care seek, understand, and apply evidence will follow in Chapter 5. Meanwhile, the next section presents the findings from the second half of this thesis, which is an analysis of the effectiveness dimension of the HTA framework (Figure 3.2). The methodologies of systematic review and critical appraisal were utilized in order to examine the current state of knowledge in regard to Lovaas Autism Treatment.
Analysis of Lovaas Effectiveness

Search Results & Study Selection

A systematic search of major databases identified 285 citations and abstracts referring to Lovaas therapy; there were no articles identified from an additional fugitive search of the grey literature. Searches were conducted between December 2004 and January 2005, with updates run in December 2005 and again in March 2006. Years searched were from January 1998 to March 2006. Along with the electronic search, an additional article was located through contacting colleagues and researchers in the field. In addition, a key researcher of Lovaas Autism Treatment – Dr. Tristram Smith – was contacted in January 2007 and provided with the list of retrieved studies. Dr. Smith replied with a reference list of replications and partial replications of the UCLA Model; no new studies were identified.

Two researchers (SB & CG) independently reviewed the search results, narrowing the list down to 58 articles that met the minimum inclusion criteria. Full-texts were retrieved and the criteria re-applied in order to discern which studies met the specifications for a critical appraisal; six articles remained after the review. Two studies were randomized controlled trials (Sallows & Graupner, 2005; Smith, Groen & Wynn, 2000); and four were observational studies (Cohen, Amerine-Dickens & Smith, 2006; Eikeseth, et al., 2006 & 2002; Howard, Sparkman, Cohen, Green & Stanislaw, 2005). Figure 4.2 displays the study selection process.
Appendix M describes the reasons for exclusion for each of the 52 retrieved studies, while Table 4.1 summarizes the number of studies that met each of these exclusion categories. The six studies that met the inclusion criteria are summarized in detail in Appendix N.

Table 4.1 Summary of Reasons for Exclusion of Studies

<table>
<thead>
<tr>
<th>Exclusion code</th>
<th>Description</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECL</td>
<td>Editorial, comment, or letter</td>
<td>1</td>
</tr>
<tr>
<td>NR-design/methods</td>
<td>Not relevant: design or methodological issues</td>
<td>17</td>
</tr>
<tr>
<td>NR-IA</td>
<td>Not relevant: implementation/application issues</td>
<td>0</td>
</tr>
<tr>
<td>NR-OCD</td>
<td>Not relevant: opinion/commentary/description</td>
<td>10</td>
</tr>
<tr>
<td>NR-ROS</td>
<td>Not relevant: report of study</td>
<td>0</td>
</tr>
<tr>
<td>NR-Review</td>
<td>Not relevant: review/overview</td>
<td>17</td>
</tr>
<tr>
<td>NR-Stat Meth</td>
<td>Not relevant: statistical methodology issues</td>
<td>0</td>
</tr>
<tr>
<td>NR- Other</td>
<td>Not relevant: other reasons/issues</td>
<td>7</td>
</tr>
<tr>
<td>NR- Text word only</td>
<td>Not relevant: studies identified had text word in title and/or abstract, but not relevant content.</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>52</td>
</tr>
</tbody>
</table>

Figure 4.2 Flow Diagram of Study Selection Process

Potentially relevant citations initially identified (n=285)

Citations identified from other sources (n=1)

Total citations identified (n=286)

Studies retrieved for more detailed evaluation (n=58)

Studies excluded with reasons (n=52)

Relevant studies included in systematic review (n=6)
Study Quality Assessment

Observational Studies: Reisch Quality Assessment Tool

The quality of the observational studies was assessed by applying Reisch, et al.'s (1989) Check List for Assessing Therapeutic Trials (Appendix H). The Reisch tool outlines the criteria that its authors have identified as important for determining a study's validity. In particular, 33 of the criteria were deemed "primary criteria", meaning that they were considered essential components of a solid research design. These primary criteria were “starred" with an asterisk (*) on the Check List for easy identification. Each criterion was phrased as a question requiring either a yes, no, unclear/unknown or not applicable response. Criteria were considered not applicable “only when they could not be applied if the study were properly designed to answer the question being investigated." (p. 815).

Upon applying the tool to the four retrieved observational studies, it was determined that Section 13 of the Check List: "Evaluation of Subjects and Treatment/Management", criteria F: "Evaluation of potential hazards", did not apply to Lovaas therapy, therefore, this section was scored as N/A. However, since this item contained one of the identified "primary criteria", the number of criteria for this section was reduced from 7 to 6. This resulted in a new total of 32 primary criteria for the entire tool.

Next, each criterion was systematically applied to the study report with its presence/absence determined and marked on the Check List. Once all the criteria were assessed, the final step was to rate the paper. Rating 1 was a subjective rating whereby the reviewer assigns a score to a criterion based on how well they ascertained the researchers did on that particular item. A score of 2 indicated superior performance, 1 meant undistinguished performance, while a rating of 0 denoted poor performance. The scores for each quality indicator of Rating 1 were then summed and divided by the maximum possible total, resulting in a final subjective index of the study's quality. Conversely, Rating 2 was an objective approach that calculated an index of the ratio of primary criteria fulfilled to the total number of primary criteria that were applicable. This process consisted of summing the number of starred primary criteria identified in the study and dividing by the maximum possible total. The result was an objective index of the study's quality.

Table 4.2 Reisch Quality Assessment Indices for the Included Observational Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Rating 1* (Subjective)</th>
<th>Rating 2* (Objective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen et al., 2006</td>
<td>0.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Eikeseth et al., 2002</td>
<td>0.36</td>
<td>0.50</td>
</tr>
<tr>
<td>Eikeseth et al., in press</td>
<td>0.36</td>
<td>0.50</td>
</tr>
<tr>
<td>Howard et al., 2005</td>
<td>0.11</td>
<td>0.22</td>
</tr>
</tbody>
</table>

* Ratio scores
Based on the above assessment, the reviewers determined that the four observational studies were all of poor quality, and two of the six studies did not meet all the inclusion criteria (Sallows & Graupner, 2005; Smith et al., 2000). Under normal circumstances these would have been excluded from a review. However, based on the paucity of research in this field, it was agreed that a review would proceed, as any information gained from analyzing this work would be beneficial to future researchers.

**Randomized Controlled Trials: Modified Jadad Quality Assessment Tool**

The two RCTs were assessed utilizing the *Modified Jadad Tool* (Appendix G). As previously described in the Methodology chapter of this thesis, this tool was further modified to fit the design of behavioural research by omitting the reference to adverse events and re-wording the criterion pertaining to double-blinding to refer to blinding of the outcome assessor. Again, each of the five items were systematically applied against the retrieved studies taking into consideration the Guidelines for Assessment that were included with Jadad’s tool (however, the tool did not include suggestions for the additional criteria of “inclusion/exclusion” and “statistical analysis”). Table 4.3 summarizes the quality assessment scores for the two retrieved RCTs.

**Table 4.3  Jadad Quality Assessment Indices for the Included Randomized Controlled Trials**

<table>
<thead>
<tr>
<th>Item</th>
<th>Sallows &amp; Graupner, 2005</th>
<th>Smith, Groen &amp; Wynn, 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomization</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Blinded Assessor</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Inclusion/Exclusion Criteria</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Statistical Analysis</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Withdrawals/Dropouts</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>SCORE</strong></td>
<td><strong>4/7</strong></td>
<td><strong>6/7</strong></td>
</tr>
</tbody>
</table>

As described in the Methodology chapter, the Jadad tool assigns additional points if a criterion is adequately implemented. However, specific to Sallows & Graupner (2005), no additional points were awarded for the randomization criterion as randomization was not successful (it resulted in unequal groups). For the Smith, Groen & Wynn (2000) study, an additional point was awarded for the randomization criterion since the paper described an independent statistician pairing the children based on IQ, then utilizing a random numbers table to assign one member of each pair to either treatment groups. However, the study did not warrant a bonus point for the blinded assessor criterion because the assessor that administered the Early Learning Measure was not blinded. (Although the authors attempted to control for any
potential bias by videotaping sessions, having them scored by a blinded independent rater, then determining the inter-rater reliability).

Based on the above scores, it was determined by both reviewers that the studies were of acceptable methodological quality and were to be included in the systematic review.

**Critical Appraisal of Study Methods**

As previously described, the three objectives of critical appraisal are, (a) to understand the rigor of the studies, (b) to uncover reasons for differences among study results; and, (c) to provide readers with the required information to judge the applicability of the review to their clinical practice (Meade & Richardson, 1997). The studies were critically appraised through the application of six criteria identified by the American Autism Working Group as essential diagnostic, methodological, and statistical issues that should be addressed in research designs of behavioural and social intervention research (Bristol, et al., 1996):

1. The study should compare various approaches to treatment.
2. The study should involve random assignment to different treatment conditions.
3. The study should use standard intervention protocols that capture a wide range of skills and symptoms, under both laboratory and “real life” situations.
4. The study should make use of outside evaluators who are not invested in the outcome of the research.
5. The study should assure high compliance with the defined treatment protocol to ensure that the intervention was actually and consistently implemented as designed.
6. The study should use longitudinal designs that evaluate treatment effects, both during the treatment itself, and at set points after the intervention has been accomplished.

**Cohen, et al., 2006**

**Criterion #1: The study should compare various approaches to treatment.** This study (part of the Central Valley Autism Project in Modesto, California) met this criterion as it compared “Early Intensive Behavioural Treatment” (EIBT) to a “Comparison Group”. The EIBT group received Lovaas therapy without the use of aversives, while the comparison group participated in a variety of community services (although not well defined) selected by the child’s parents (similar to a “standard of care” arm).

**Criterion #2: The study should involve random assignment to different treatment conditions.** The objective of this criterion was to assess the study for possible subject selection biases that would render the groups dissimilar. The Cohen report did not meet this standard as treatment allocation was not concealed nor were the children randomized. As described in Appendix N, the authors claimed it was both illegal and unethical to randomize to treatment due to the mandate of their community program in which their research is based. Instead, parents were allowed to choose their child’s treatment regime from a menu of options (which
included EIBT); if there was no space in the EIBT program (due to a shortage of therapists), this option was not offered. If this was how the children were assigned to treatment conditions each child would have had an unequal chance of receiving the intervention. However, upon closer scrutiny of the study, there is a question as to whether this is a prospective design or a convenience sample based on the administrative records of existing children from the CVAP agency. It is possible that their standardized annualized scores had been analyzed retrospectively as there was no mention as to whether the EIBT group had been enrolled first then followed prospectively. However, they stated that the comparison group, which was assembled through a file review and matched to an EIBT child based on age and IQ, was followed prospectively. Unfortunately, the method for matching was not very explicit in regard to its criteria, thus the study was subject to selection and allocation biases (systematic differences in comparison groups). In fact, the results showed that the matching was not successful as there are significant differences between the groups at intake on four variables: (a) diagnosis, (b) mother’s level of education, (c) father’s level of education, and (d) two-parent households. In addition, the age of the subjects at intake was not provided.

In an attempt to control for confounders, statistical measures were employed, however the results showed that differences in outcomes between groups were still present. One of the variables that could not be ruled out as a potential confounder of IQ was father’s education. In addition, the authors also acknowledged that since there were more children with the diagnosis of PDD-NOS in the comparison group it might have biased the results in its favour. This is another indication of differences between groups that could have accounted for the difference in results. Any differences between groups appeared to be strictly due to inadequately controlled groups being compared.

Another limitation of the study was that the authors did not report how the children were recruited. Based on the statement that funding had been divided between the Valley Mountain Regional Centre (VMRC) and the child’s Special Education Local Planning Area (SELPA), it is likely that the children were referred to the centre by VMRC. However, without additional information (e.g., whether the referring physicians had prior knowledge of the study, or whether all children who were referred were offered the opportunity to participate), it is difficult to determine whether another possible source of selection bias – referral bias - had occurred.

Sample size was not pre-determined; instead a “recruitment period” from 1995 to 2000 was reported. If this was a prospective design, children who were entered into the study in 2000 would have had to been followed for the full three years to 2003. Hence, the investigative team would have been involved in the study for a minimum of eight years; this was not reported as such. In addition, one of the inclusion criteria stipulated that a child could not have received more than 400 hours of behavioural intervention prior to intake. Presumably anything more could have potentially confounded the results, although no support for this assumption was provided.
In light of the preceding comments, this inclusion criterion also supported the hypothesis of a retrospective study design.

**Criterion #3: The study should use standard intervention protocols that capture a wide range of skills and symptoms, under both laboratory and "real life" situations.** This particular criterion addressed the issue of performance (or, intervention) bias. Performance bias refers to "systematic differences in the care provided to the participants in the comparison groups other than the intervention under investigation" (Alderson et al., 2004, p. 52). In particular, a study must be protected against contamination (the provision of the intervention to the control group) and co-intervention (the provision of unintended additional care to either comparison group) (p. 52). In the Cohen study, the EIBT group was reported to have received between 35-40 hours per week of treatment. The intervention consisted of: (a) in-home 1:1 instruction, (b) peer play training, and, (c) regular education classroom inclusion. The design also incorporated monitoring of pre-determined developmental markers at 6, 12, 24 and 36 months. These indicators were used to assess whether treatment should continue for a particular child. The number of hours of in-home instruction was based on the age of the child; those older than 3 years of age received 35-40 hours per week and children less than 3 years old were administered 20-30 hours per week. After approximately one year, the number of home hours were reduced and replaced with an increase in peer play and classroom instruction. A trained tutor facilitated peer play sessions that began with the introduction of one typically developing peer and progressed up to group play once 90% of the criteria for interacting with peers were met. Concurrently, the EIBT children also entered a teacher-directed structured regular education preschool setting. The tutor accompanied the child into the classroom, not to provide 1:1 instruction, but in the role of a classroom aid. As the children progressed, treatment hours were gradually reduced to zero, although family and school consultation continued for 1 to 2 hours per month for an additional 1-2 years.

The comparison group received therapy based on the parent's choice from a selection of options. As treatment options were eclectic, no standardized procedures or measurements of treatment fidelity were implemented. Although not specifically identified as such, this group would be considered the standard of care arm of the study. The authors reported that one child in the comparison group received less than 9 hours of discrete trial training per week, and that in total, 17 children were enrolled in Special Day Classes (SDC) of varying intensity which included up to five hours per week of "behavioural therapy" in their curriculum. It is uncertain whether this intensity of behavioural therapy constituted treatment contamination.

In the third year of the study the EIBT group deviated slightly from the original Lovaas/UCLA model when "advanced social skills training" was added to their curriculum. The researchers felt that "many children" (exact numbers not reported) had made significant gains during the first two years of intervention and thus required additional training in social skills.
beyond the UCLA curriculum. However, the authors recognized that this might have confounded Year 3 outcomes. Unfortunately, because so little is known about the treatment protocol for the comparison group, this critical appraisal criterion of standardized protocols was not satisfied.

**Criterion #4: The study should make use of outside evaluators who are not invested in the outcome of the research.** The objective of this criterion was to assess for measurement (detection) bias. Detection bias refers to “systematic differences between the comparison groups in outcome assessment” (Alderson et al., 2004, p. 52). This criterion was not met. First, the researchers used a licensed psychologist who was independent of the study and reportedly blinded to treatment assignment to conduct the intake assessments and to make a DSM-IV diagnosis. Next, each child's diagnosis was confirmed by a certified examiner under the employ of the research site utilizing the Autism Diagnostic Interview Revised (ADR-R) assessment tool. It is uncertain as to whether this examiner was blinded to group assignment or if there was any disagreement over diagnoses (and if so, how they were resolved). The children were then assessed throughout the study by “an independent, self-employed, highly-skilled, licensed, child evaluator” (p. S149) employed by the research site; the researchers noted that this evaluator only received the child's name, birth date, parent's names and telephone number. However, the authors stated that they felt this was not the best assessment protocol because they were unable to institute measures to ensure both blinding of treatment and reliability of test administration and scoring. Although it might have been difficult to ensure blinding of treatment (as parents could have inadvertently mentioned some aspect of their child's treatment program that would un-blind treatment assignment), I feel that intra-rater testing could have been incorporated into the assessment protocol to determine the reliability of testing procedures. And finally, when attempting to control for potentially confounding demographic variables, the researchers reported that due to the small sample size, their ANCOVA analysis might not have yielded reliable results.

**Criterion #5: The study should assure high compliance with the defined treatment protocol to ensure that the intervention was actually and consistently implemented as designed.** Once again, this criterion was looking for performance (intervention) bias in order to determine internal validity of the study. Treatment fidelity (also known as treatment integrity) is the strict adherence to the prescribed treatment protocol in order to avoid treatment contamination. The authors pointed out that it was impossible to assess treatment fidelity of the comparison group because of the diverse programs involved. Therefore this quality assessment criterion could only be applied to the EIBT group. Here, staff were trained and supervised by UCLA consultants (as described in Appendix N). The authors incorporated an additional quality assurance process into the research design by measuring the tutors' adherence to UCLA procedures. This consisted of videotaping 12 tutors and having blind raters score their
performance. Although no information is provided as to the specific criteria measured, the authors stated that their adherence to protocol was found to be "non-significantly higher than adherence by tutors employed at UCLA" (p. 5148). However, due to the inability to measure treatment fidelity in the comparison group (and possibly the EIBT group as well), this criterion did not appear to have been met.

**Criterion #6: The study should use longitudinal designs that evaluate treatment effects, both during the treatment itself, and at set points after the intervention has been accomplished.**

This criterion assessed how treatment effects were evaluated and measured, with the intent of identifying any possible measurement or classification biases. As previously described the study design of Cohen et al. is in question. Initially the study appeared to be a prospective design of three years duration with evaluations of recognized intellectual, mental, social, language, and behavioural indicators administered annually to both groups. However, mention is only given to the comparison group being followed prospectively. If this were the case, the two groups would have been evaluated at different times during the possible eight years. As previously mentioned, outcome assessment was attempted in a blinded fashion, but, as admitted by the authors, this could not be assured. Therefore, this criterion was not met.

IQ is the primary outcome measure. Both groups received the same testing pre and post treatment. Although the authors did not mention the validity of the tools used, my critical appraisal of the Eikeseth et al. studies (to follow) points out that, although frequently used, the Reynell and the Vineland have not been validated in an autistic population.

Some missing data in Cohen's study made the interpretation of treatment effects difficult. For example, the intake means and standard deviations of the chronological ages of the children, along with the number of hours of previous EIBT treatment, might have been important variables to compare and incorporate into the statistical control (if significant differences between groups had been discovered). Similarly, at follow-up, data on the average duration of treatment for each group, is not reported. The eclectic nature of the treatment options in the comparison group resulted in large variances in treatment intensity (1-4 hours per week for an in-home developmental intervention; 9 hours per week for the Early Start Autism Intervention Program; and, between 15-25 hours per week for special education classes); however, since there has been much debate about whether observed treatment effects of EIBT are actually due to the therapy, as opposed to the amount of attention a child receives (Dawson & Osterling, 1997), controlling for intensity of treatment might have contributed useful data to this discussion.

The sample size of both groups changed throughout the study. Originally, the EIBT group contained 24 participants while the comparison group had 23; however this was later reduced to 21 for each group. These five dropouts were not included in the data analyses. (Explanations for dropping out along with the comparability of characteristics between dropouts and those
that participated were provided). Later, four more children from the EIBT group withdrew early (one was able to integrate into a regular classroom after two years; the other three were failing to meet pre-determined developmental markers and were withdrawn). However, all these children completed follow-up assessments, and were included in the statistical analyses. There were no further dropouts reported for the comparison group.

**Summary of critical appraisal of Cohen et al., 2006.** The purpose of Cohen's study was to replicate the original Lovaas/UCLA model within a community setting. However, methodological limitations are numerous with the major ones being the ambiguity of the research design (and hence, lack of randomization) and post-hoc analyses. In the preface to this study published in *Developmental and Behavioural Pediatrics*, although the journal editor commends Cohen and colleagues on the over-all design of the study, the lack of randomization was identified as a significant limitation. While it was previously mentioned that the EIBT group had more children with PDD-NOS (and therefore, possibly better outcomes because of this), another confounder could have been the number of hours of behavioural intervention received prior to intake (mean treatment hours were not reported). In light of Sallows & Graupner's (2005) study (review to follow), which found a distinction between rapid learners and moderate learners, previous behavioural treatment might have been another significant variable to control as children who had received 400 hours of intervention may have had an advantage to those who had not received any prior treatment.

When interpreting outcomes, the above considerations must be taken into account. Although the researchers attempted to control for differences in intake variables, they acknowledged that interpretation of results should proceed with caution due to the large number of variables relative to the small sample size. The authors also recognized that the addition of advanced social skills training might have made it difficult to compare the third year data. However, the treatment did not make any of the children any worse. None of the outcome measures indicated that the application of EIBT resulted in regression.

Of interest is that Lovaas cited this study in a letter to the editor of *Time* magazine rebutting an article they had published on autism (Lovaas, 2006).

Overall, Cohen et al. only satisfies one of the six critical appraisal criteria (#1). Therefore, chance differences make it difficult to interpret this study.

**Eikeseth et al., 2002 & in press**

**Criterion #1: The study should compare various approaches to treatment.** Eikeseth's studies met this criterion in that the researchers compared intensive "behavioural treatment" against intensive "eclectic treatment" for children who began treatment later in life – between 4 and 7 years of age. The authors described behavioural treatment as being based on the UCLA model (without the use of aversives) administered for a minimum of 20 hours per week. The eclectic treatment is reported as a variety of different interventions that were reflective of best
practices, also delivered at a minimum of 20 hours per week. This arm of the study was also similar to a standard of care protocol. Both groups were treated within the public school system. Eikeseth et al. (2002) followed the children for one year, with Eikeseth et al. (in press) following the groups forward to the age of eight years.

**Criterion #2: The study should involve random assignment to different treatment conditions.** This criterion was not met. In Eikeseth's study, instead of being randomly assigned to treatment groups, participants were allocated by an independent clinician based on the availability of a behavioural treatment supervisor. Only a cursory explanation of the allocation process is provided. All referrals to the centre between November 1995 and November 1998 were recruited into the study. While ensuring that all referrals are included may eliminate one of the concerns of referral bias, this specific type of selection bias may still be a possibility since the children who were referred to the centre may have been different from those that were not referred (it is not known whether the referring physician(s) had any knowledge of the ongoing study). Therefore, referral bias could have impacted the generalizability of the study, as these children may not have been representative of the defined population.

The lack of randomization resulted in differences between groups in regard to an imbalanced sex ratio. The behavioural group consisted of 13 children, of which 5 were girls, while the eclectic group had 12 children, which included 1 girl. (Girls are known to do worse than boys). While the authors purported that the groups were not significantly different on any of the intake variables, they also pointed out that the eclectic group attained higher average scores than the behavioural group on 10 of the 11 intake variables, thus surmising that the eclectic group could have been functioning at a higher level.

Eikeseth's study outlined three inclusion criteria: (a) a diagnosis of childhood autism as defined by the World Health Organization’s International Classification of Diseases (ICD-10), (b) chronological age between 4 and 7 years at the time of intake, (c) IQ of 50 or greater, and (d) absence of other medical conditions that could interfere with treatment. First, in regard to diagnostic criteria, Lovaas' 1987 study (and many others since) utilized the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) definition of autism, which is more popular within North America (European sites favour the ICD-10 criteria). While both diagnostic classification systems have evolved into fairly similar tools, the criteria for the DSM-IV are less strict than the ICD-10, therefore it may be more inclusive (Tidmarsh & Volkmar, 2003).

Second, the authors satisfactorily explained their rationale for the age range criterion. In fact, recruiting autistic children between the ages of 4 and 7 probably more accurately reflects the real world experience where diagnostic and treatment delays can be typical occurrences. And, third, the authors rationalize their IQ criterion of >50 by citing previous studies that indicated intensive behavioural intervention may not be efficacious below this point.
Another observation is that the authors did not proffer a rationale for the three-year recruitment period. However, they did recognize the limitation of the small sample size: "the small number of children in each of the two groups may have yielded too little statistical power to detect differences between groups" (p. 65). In addition, the criterion that the child could not have other medical conditions "that could interfere with treatment" was subjective and could have contributed to differences in outcomes.

**Criterion #3: The study should use standard intervention protocols that capture a wide range of skills and symptoms, under both laboratory and "real life" situations.** This criterion was not met. Although the authors reported that the behavioural treatment group followed the 1981 Lovaas manual and accompanying videotapes, the intervention was not adequately described. The eclectic treatment group received a variety of interventions that were individually selected for the child, such as Project TEACCH, sensory-motor therapies, in addition to the option of receiving some Lovaas Autism Treatment. The treatment plans for the eclectic group were based on recommendations from a "multidisciplinary team of school personnel" (2002, p. 55). Both groups received similar amounts of treatment (behavioural: 28 hrs/week; eclectic: 29.08 hrs/week) by therapists with equal levels of education. By age six, the intensity had been reduced to 18 hours for the behavioural treatment group and 16 hours for the eclectic treatment group. Of interest though is that the intervention protocol called for a minimum of 20 hours of intensive therapy, however, if only 20 hours had been achieved should this intervention still have been considered "intensive"?

While the original Lovaas study took place within a clinical setting, Eikeseth's research was situated within a public school with both groups attending regular classes. Total duration of treatment for both groups was one year for Eikeseth 2002. For Eikeseth (in press), the behavioural group had received 31.4 months by follow-up at age 8, while the eclectic group had received 33.3 months. As mentioned, Lovaas Autism Treatment was also part of the eclectic treatment. In fact, 42% of the eclectic group received some ABA therapy. This may be viewed as treatment contamination and could have resulted in improved outcomes for the eclectic group.

**Criterion #4: The study should make use of outside evaluators who are not invested in the outcome of the research.** This criterion was partially met as the diagnosis of autism at intake was made by an independent child psychologist (and verified by the ADI-R). However, as previously discussed, this study utilized ICD-10 diagnostic criteria as opposed to DSM-IV. Follow-up assessments were conducted by either this psychologist or an examiner with a master's degree in special education who was also independent of the study. Neither assessor had knowledge of the child's group assignment. However, procedures for ensuring blinding of the evaluators were not reported. There was one exception to the blinded outcome assessment and that was the administration of the Achenbach Child Behaviour Checklist, which was administered by teachers in an un-blinded fashion when the child reached 8 years of age.
Criterion #5: The study should assure high compliance with the defined treatment protocol to ensure that the intervention was actually and consistently implemented as designed. Both of the Eikeseth studies failed to meet this criterion. Although each group reportedly followed intervention and staff training protocols, no formal measurement of fidelity was reported. Therapists in the behavioural treatment group apparently received 10 hours per week of supervision from staff trained in the UCLA model. These supervisors subsequently met weekly with the project directors who were psychologists with 10 years of experience with Lovaas therapy. In addition, parents received training in behavioural therapy by working alongside the therapists in the schools for the first three months for a minimum of 4 hours per week. Afterwards, parents were re-directed to the home and community to implement generalization and maintenance programs. Weekly two-hour meetings were held for the behavioural treatment group whereby the child’s progress was reviewed and treatment modified if required. The therapists and parents also received additional training at that time.

The eclectic treatment group received therapy in the same one-to-one format as the behavioural treatment group; however, their therapists were not supervised. Weekly meetings were held but it was uncertain whether treatment protocols were modified to the child’s needs. Also, parent participation was not part of eclectic treatment.

Criterion #6: The study should use longitudinal designs that evaluate treatment effects, both during the treatment itself, and at set points after the intervention has been accomplished. Both of Eikeseth’s studies partially met this criterion in that, while they employed a longitudinal design and assessment of treatment effects, there were some inconsistencies with their reported outcomes. With Eikeseth’s 2002 study, treatment measures were taken after a pre-defined treatment duration of one year of therapy (“1 year follow-up”). However, in Eikeseth et al’s 2006 follow-up, (referred to as the “CA-8 follow-up” [chronological age]), all children were assessed concurrently at a specific time point. This resulted in slightly different treatment durations for each group (behavioural: 31.4 months; eclectic: 33.3 months). The authors did not report whether this was a significant difference. The treatment dosage (# of hours of therapy per week) was similar in both groups with the behavioural treatment group initially receiving 28 hours per week, while the eclectic treatment group received 29.08 hours per week for the first year and 29 hours per week up to the CA-8 follow-up. When the children entered school, the treatment levels dropped to 18 hours and 16 hours for the behavioural and eclectic groups respectively.

Measurements consisted of tests of intelligence, visual-spatial skills, language, and adaptive functioning. The authors pointed out that some of the limitations of the study were its use of measures that focused more on cognitive rather than social development and its lack of direct quality control measures of treatment. While most tests appeared to have been validated in the autistic population, the authors acknowledged that the Reynell Developmental Language Scales, although frequently used with autistic children, had not undergone research on the
reliability of its psychometric properties within this population. The authors also noted that the Vineland was “widely regarded as the instrument of choice for assessing adaptive functioning in children with autism” (2002, p. 57), however, they did not mention if the tool had been validated. In fact, the authors acknowledged that, “there [was] some evidence that Vineland scores are lower in Norwegian samples [where this research was conducted] than in the United States” (2006, p. 10). Therefore, the scores may have been underestimating these children’s adaptive functioning. All of the assessments that were conducted at screening were repeated at both follow-up time points. However, two additional measures were added after the study was initiated. The first one was the Vineland Maladaptive Behaviour Scale, which was assessed at both the 1-year and CA-8 time points, and the Achenbach Child Behaviour Checklist, which was administered only at the CA-8 follow-up.

While the results of these two studies are summarized in Appendix N, a few outcome measures deserve further scrutiny. At the 1-year follow-up, the behavioural group scored higher than the eclectic group on all measures except the Vineland Socialization, however, none of the differences on individual measures were statistically significant.

Another reported finding is that the behavioural group had less disruptive behaviour than the eclectic group at both 1-year and CA-8 follow-ups (as evidenced by the Vineland scores). This sub scale was administered to children ages 5 years and older, and since many of the children were under 5 at intake, the researchers only measured the children at the two follow-up time points. Therefore, no baseline measure was available to judge whether either group had indeed changed since intake (or whether the behavioural group always had less disruptive behaviour). However, since the mean age of the behavioural and eclectic groups at intake was 66.3 and 65 months respectively, I wonder whether some children could have been measured in order to provide this pre-treatment data. What the authors failed to report was that although the behavioural group consistently scored lower than the eclectic group (indicating less aberrant behaviour), both groups actually increased their scores from the 1-year follow-up to the CA-8 follow-up. (behavioural treatment: 4.29 to 6.3; eclectic treatment: 7.25 to 11.0).

In regard to the outcomes reported for the Achenbach Child Behaviour Checklist at the CA-8 follow-up, the authors reported group differences in social emotional functioning in favour of the behavioural treatment group on the measures of social problems and aggressive behaviour. However, as previously mentioned, as there were no pre-treatment measures taken, it can not be ascertained whether this difference was due to behavioural treatment or whether the behavioural group displayed less social and aggressive tendencies from the outset. In addition, this assessment was conducted in an un-blinded fashion by the child’s teacher.

**Summary of critical appraisal of Elkeseth et al. 2002 & in press.** The purpose of these studies was to compare behavioural treatment against eclectic treatment at the same intensity,
within the same settings, but with an older cohort of 4-7 year olds. However, there appears to be some discrepancies and missing data from the study report:

1. In the discussion section of Eikeseth 2002 (page 63), the authors noted that the behavioural treatment group gained 13 points in language comprehension; however, a quick calculation of the data presented in Table 2 on page 60 would indicate the actual gain was only 9.44 points. In addition, my calculation of the same measure for the eclectic group shows a -2.83 change (as opposed to the reported -1.0 change). In fact, all the mean changes for Reynell scores for both groups were calculated incorrectly for the comprehension, expressive and total language measures. (The primary author was contacted in regard to this discrepancy, but no response was received).

2. The study's section: Achievement of Scores within the Average Range (p. 9) described gains made by 54% (7 of 13) of the children in the behavioural group and 17% (2 of 12) of the children in the eclectic group. However, the authors did not include a table to support their claims. In addition, gains/losses experienced by these children in regard to their Vineland Maladaptive Behaviour and Achenbach Child Behaviour Checklist scores were once again referenced although no pre-treatment measure was available in which to compare.

3. In the Discussion section of the "in press" report, the researchers stated that "Vineland Composite scores increased throughout ABA treatment" (p. 13), however, by pulling in the data from the 2002 study, the behavioural treatment group's scores were 55.77 at intake, 67.0 at 1 year, and 67.9 at CA-8 follow-up, indicating a levelling off after the first year.

4. Between intake and CA-8 the authors claimed the eclectic treatment group dropped 12 points on the Vineland Socialization scale (in press, p. 12). However, a comparison of the data from Table 2 in Eikeseth (2002) shows the intake score to be 62.17, while Table 2 in Eikeseth (in press) documents the CA-8 score to be 58.1; a drop of only 4.07 points.

5. 8-year-old follow-up assessment data were missing the follow-up Reynell scores. It is uncertain whether these were measured.

Eikeseth only satisfies one of the six critical appraisal criteria (#1), and partially meets criteria #2, 3 and 5. Once again, a lack of randomization weakened this study.

Howard, et al., 2005

Criterion #1: The study should compare various approaches to treatment. This criterion was met as Howard et al. compared "intensive behavioural treatment" (IBT group) against both "intensive eclectic treatment" (Autism program - AP group) and "generic early intervention" (Generic program - GP group).
Criterion #2: The study should involve random assignment to different treatment conditions. This criterion was not met as children were not randomized to treatments nor was treatment allocation concealed. The authors reported that educational placement decisions were made by public school administrators (with input from parents); however, no further information is proffered as to how these decisions were made. At intake, the authors stated that all three groups were similar with the only exception being that the nonverbal skills scores were higher for the generic intervention group. At first glance, an examination of the table of participant characteristics appears to show otherwise. First, both the intensive behavioural group and the generic program group have a higher ratio of boys to girls than what would be expected in the normal population. Second, sample size is also smaller for the AP and GP groups (n=16 in each comparison group, compared to n=29 in the IBT group). Third, the IBT group is more likely to have two Caucasian parents. Fourth, the percentage of children with AD and PDD-NOS varies both within-groups and between-groups. And, fifth the marital status of the child's parents also varies with fewer married parents in the generic intervention group. However, the researchers stated that none of these differences between groups were statistically significant. They did however acknowledge that the age at diagnosis and parental education differed between the groups therefore they attempted statistical control in their analysis to compensate.

Children were recruited from referrals from non-profit agencies between 1996 and 2003. Again, it was unknown whether referring physicians were aware of the study. Justification for the sample size was also not provided. In addition, the inclusion criteria did not stipulate a minimum entrance IQ level. At intake the IBT group was reported to have a mean IQ of 58, the AP group 54, and the GP group 60, all of which were statistically non-significant.

Criterion #3: The study should use standard intervention protocols that capture a wide range of skills and symptoms, under both laboratory and 'real life' situations. This criterion was not met. In the intensive behaviour analytic treatment group (IBT), the children received between 50-100 "learning opportunities" per hour based on discrete trial training (DTT), incidental teaching and "other behaviour analytic procedures." The researchers noted that "programs similar to those described in several treatment manuals... were delivered using a combination of behaviour analytic techniques." Yet, they do not explicitly identify the programs (instead, they cite references). This made it difficult to judge treatment fidelity to Lovaas therapy; therefore, I classified the intervention as "Lovaas-type" therapy. The intensity of the program was slated to be between 25 and 40 hours (based on age), but actual hours of therapy received were not reported.

Children in the Intensive Eclectic (AP) comparison group received an assortment of behavioural treatments (including DTT). Although the intensity of the program (25-30 hours/week), the therapist-to-student ratio (1:1 or 1:2), the site (public special education
classrooms, and staffing (teachers & aids) were all described, no information was provided on the actual amount of treatment hours received or the combination of therapies utilized within this program.

And finally, the authors acknowledged that the generic, non-intensive early intervention program (GP) was not operationally defined and actual treatment hours received were not documented.

**Criterion #4:** The study should make use of outside evaluators who are not invested in the outcome of the research. Howard did not meet this criterion. While independent examiners utilizing DSM-IV criteria made the diagnosis of autism or PDD-NOS, outcome assessors (who were described as being “independent”, and “not involved in delivering treatment”) were un-blinded to treatment assignment.

**Criterion #5:** The study should assure high compliance with the defined treatment protocol to ensure that the intervention was actually and consistently implemented as designed. This criterion was not met. Although parents and therapists in the intensive behavioural group received training and consultations with senior therapists, actual hours of treatment received were not reported. While attempts were also made to ensure treatment integrity (e.g., direct observation, videotaping, feedback), formal measurements were not conducted for any of the groups. Therefore, internal validity cannot be assured.

**Criterion #6:** The study should use longitudinal designs that evaluate treatment effects, both during the treatment itself, and at set points after the intervention has been accomplished. This criterion was not met. The study design called for follow-up assessments conducted after approximately 14 months of treatment; however, the authors did not offer a reason for the selection of this time point. Data were described as “interim”, so future follow-up appears imminent. Standardized measures of cognitive, language and adaptive skills were assessed pre- and post- treatment on all groups in an un-blinded fashion. This created a potential bias in reporting treatment effects (however, the direction of the bias would be difficult to determine). The researchers acknowledged that the short length of the study might have limited the magnitude of the outcomes.

Drop out rates were high for all groups. Originally, 37 children met the eligibility criteria for the IBT group; however, only 29 were included in the final analysis fourteen months later. For the AP and GP groups, 41 children met the criteria, but only 32 (16 in each group) were included. Reasons for withdrawals were reported and included issues such as parents being unable to establish intensive programs in their homes, families moving away, inability to reach parents despite repeated attempts, and refusal for follow-up testing. In addition, two of the children in the IBT group were withdrawn because attempts to increase the intensity of their program resulted in increased behavioural problems.
Summary of critical appraisal of Howard et al., 2005. The purpose of this study was to compare intensive behavioural analytic intervention (IBT) against two controls— an intensive eclectic (AP), and a non-intensive generic early intervention program (GP). However, since no between-group differences were discovered at follow-up between the AP and GP groups, their data were combined (which shows statistically significant differences). The authors reported that the IBT group received between 25-40 hours of therapy per week, however, they did not provide a mean so that comparisons could be made relative to the intensity of the eclectic group (30 hours), and the generic program (15 hours). For example, if the IBT mean was closer to 40, it would be difficult to discern whether the observed between group difference was due to the intensity of the program as opposed to the actual treatment delivered.

Lovaas also cited this study in his response to the Time magazine article previously mentioned (Lovaas, 2006).

Howard et al. only satisfies one of the six critical appraisal criteria (#1), thus making it a weak study design.

Sallows & Graupner, 2005

Criterion #1: The study should compare various approaches to treatment. This criterion was not met, as both groups received Lovaas treatment (delivered by different providers and at different intensities). The aim of this study was to replicate the UCLA model in a community setting (as opposed to an academic setting) and to compare the results to the outcomes of the original Lovaas study (1987) and to the control group who received a less costly parent-directed program of Lovaas therapy.

Criterion #2: The study should involve random assignment to different treatment conditions. This criterion was not met. Although the authors reported that the children were recruited from local special education programs, matched on pre-treatment IQ scores and then randomly assigned by an independent statistician to either the clinic-directed group (UCLA intensive behavioural treatment) or the parent-directed group (less intensive UCLA model), the differences between groups at intake attests to the fact that randomization failed. When groups were compared, a pattern emerged whereby changes in scores on IQ showed that the children made either rapid progress or moderate progress, therefore, their data were analyzed separately. When characteristics of “rapid learners” (post treatment IQs >85) and “moderate learners” (post treatment IQs <85) were examined, rapid learners in the clinic-directed group had higher pre-treatment IQs, Vineland scores, and levels of verbal imitation.

Criterion #3: The study should use standard intervention protocols that capture a wide range of skills and symptoms, under both laboratory and ‘real life’ situations. This criterion was met. The authors noted that the treatment procedure for both groups was based on Lovaas’ curriculum (without aversives) “with the addition of procedures supported by subsequent research. . . . which have been widely disseminated [italics added].” (p. 422). It is unknown
whether these new procedures would have altered the treatment intervention enough to dispel the claim of a Lovaas replication or if it resulted in intervention bias from the addition of co-interventions. Strict protocols were in place for staff training and measurement of treatment fidelity. However, 22 of the 23 children also received some form of supplemental treatment either prior to the study or during the first year. This included therapies such as special education, preschool, speech, sensory integration, auditory integration, music therapy and horseback riding. The researchers determined that these correlations were low and non-significant.

Treatment intensity goals were met for the clinic-directed group with the children receiving 39 hours per week for the first year and 37 hours per week for the second year. However, treatment intensity for the parent-directed group surpassed the researcher’s expectations with the children receiving 32 hours per week for the first year and 31 hours per week for the second year. Because of the parents diligence in administering the program, between-group differences based on intensity were difficult to assess. Instead, children were analyzed as “rapid learners” and “moderate learners” as previously described.

**Criterion #4:** The study should make use of outside evaluators who are not invested in the outcome of the research. This criterion was only partially met. Diagnosis was made by an independent psychologist based on DSM-IV criteria. Pre-treatment assessments were conducted by the second author of the study prior to treatment assignment. To counter possible claims of bias, an independent psychologist also tested the IQs of one quarter of the children with non-significant differences noted. Follow-up assessments were conducted in a blinded fashion for the “rapid learners” only; the “moderate learners” were assessed once again by the second author (to reduce cost). The authors disputed the argument of possible measurement bias, relying on the high correlations between the second author and the independent psychologist on intake IQs.

**Criterion #5:** The study should assure high compliance with the defined treatment protocol to ensure that the intervention was actually and consistently implemented as designed. This criterion was met. Treatment fidelity was monitored and recorded. The senior therapists and clinic-directed therapists had to meet quality control criteria that were originally developed from the UCLA program. This involved a written exam, followed by a videotaped review of their work. In addition, senior therapists received weekly supervision by the senior author. Progress reviews were held weekly for the clinic-directed group, and every 2 months for the parent-directed group. Treatment plans could be revised accordingly. As mentioned, known co-interventions were identified and controlled for statistically.

**Criterion #6:** The study should use longitudinal designs that evaluate treatment effects, both during the treatment itself, and at set points after the intervention has been accomplished. This criterion was also met as Sallows and Graupner utilized a four-year longitudinal design with yearly evaluations. Outcome measurements included a comprehensive battery of tests of
intelligence, language, adaptive behaviour, academic achievement, and rate of acquisition of skills. In addition, the researchers assessed the rapid learners for residual symptoms of autism along with examining variables that could be predictors of outcomes. As previously mentioned, one limitation in regard to outcome measurements was the second author assessing the moderate learners.

Summary of critical appraisal of Sallows & Graupner, 2005. This study was described as "the first replication of the UCLA Model that included all of the elements identified by Lovaas." (Cohen, et al., 2006, p. S146); in addition, it was personally acknowledged by Lovaas (Lovaas, 2006). This research met two of the six critical appraisal criteria (#3 & 5), and partially satisfied criteria #4. Since the purpose of this study was to replicate the original UCLA model in a community setting (in addition to looking at residual symptoms of autism and predictors of outcomes), it was not addressing an effectiveness question, therefore no true comparator group was included. As such, this study adds little information to this debate.

Smith, et al., 2000

Criterion #1: The study should compare various approaches to treatment. This criterion was not met because the study was not designed to compare different treatments. Instead, it looked at a group of children who received intensive Lovaas therapy ("intensive therapy"), (albeit intentionally not as intensive as the original Lovaas study), against Lovaas therapy administered by parents ("parent training"). In addition, the authors wanted to extend the research literature by enrolling not only children with autistic disorder (AD) but also children with pervasive developmental disorder not otherwise specified (PDD-NOS).

Criterion #2: The study should involve random assignment to different treatment conditions. This criterion was met. After standardized intake assessments were completed, the children were divided into cohorts based on diagnosis (AD & PDD-NOS). An independent statistician paired the children from each cohort based on pre-treatment IQ, then, utilizing a random numbers table, one member of each pair is assigned to a treatment arm. Therefore, both the allocation and the sequence of allocation were concealed from the investigators limiting allocation bias. The authors reported that any between-group differences at intake were not statistically significant, however, no statistical analysis was offered on subject characteristics ("The groups appeared [italics added] similar on all variables" (p. 272). The study's Table 1 seems to show differences between groups in regard to years of schooling for parents (with the parent-training group more highly educated). In regard to intake scores, there also appears to be a significant difference between the PDD-NOS children in the intensive treatment and the parent training groups in regard to their mean Developmental Language Scale scores (DLS). Although the authors reported no significant differences on intake, as per their Table 2, the PDD-NOS children in the intensive treatment group had a significantly lower DLS scores than the autism subgroup at follow-up.
Children were recruited from referrals to the UCLA Young Autism Project between 1989 and 1992. The authors stated that all referrals were screened for the study, however 8 children did not meet the criteria (reasons were given). The groups were fairly evenly split in regard to diagnosis and sex ratio. In addition, there were no dropouts during the course of treatment. No justification for sample size was proffered yet it was acknowledged that due to the small sample size, and skewed distributions, certain statistical tests could not be conducted.

**Criterion #3: The study should use standard intervention protocols that capture a wide range of skills and symptoms, under both laboratory and ‘real life’ situations.** This criterion was met. Both treatment groups received manualized treatments based on the UCLA program (Initially, four children were also administered “contingent aversives”, but this practice was later halted for all children). The authors attempted to simulate “real life” by acknowledging the cost of service delivery and the stress that intensive programs can place on families and children by designing an intervention protocol that was not only less intensive than the original Lovaas study, but also administered in more realistic settings (community and home). However, one difference between this replication study and Lovaas' original study is that this study had an option to withdraw children who had not met specific developmental goals after 18 months of treatment.

Due to staffing issues, treatment goals for the intensive treatment group were not achieved. Instead of 30 hours per week, the children ended up receiving only 24.52 hours per week for the first year, thereby placing the intensity of the program in jeopardy. The authors acknowledged that measures such as paying therapists (as opposed to relying on students) might have prevented this scenario. The number of hours of classroom instruction was not reported. For the parent-directed group, while the number of hours of special education classes was provided (10-15 hours per week), the actual number of treatment hours administered by the parents was not. (Based on a review of the study, this writer estimates this to be approximately 10 hours per week for the first 3-9 months, decreasing to 5 hours per week thereafter).

**Criterion #4: The study should make use of outside evaluators who are not invested in the outcome of the research.** This criterion was only partially met because although evaluators were independent of the study and blinded to treatment assignment, those that conducted the Early Learning Measure and the Achenbach & Teacher Form were not, therefore bias may have been introduced. The authors reported that although diagnoses for all the children were made by independent psychologists, no standardized diagnostic instrument was used, nor were follow-up diagnostic assessments made.

**Criterion #5: The study should assure high compliance with the defined treatment protocol to ensure that the intervention was actually and consistently implemented as designed.**

This criterion was partially met. Measures to monitor treatment fidelity were in place but not consistently applied to both groups, for instance, the recording of intervention intensity and classroom instruction. However, measures for selecting, training and supervising the student
therapists are reported as being consistent with protocols followed in the original Lovaas study. In addition, criteria were developed for therapists to progress to positions of supervisors based on theoretical exams and clinical reviews. There was no indication of the administration of co-interventions.

**Criterion #6: The study should use longitudinal designs that evaluate treatment effects, both during the treatment itself, and at set points after the intervention has been accomplished.**

This criterion was met in that the Smith study was a longitudinal design with outcome assessments conducted between the ages of 7 and 8 years. However, results showed that there was a wide range in treatment duration for the intensive intervention group (between 18.4 months to 30.79 months).

The outcome measures incorporated standardized instruments for intellectual functioning, language functioning, adaptive functioning, socio-emotional functioning, academic achievement, class placement, progress in treatment and parent evaluation. The other outcome measures were described as "standardized instruments", however as previously mentioned, the psychometric properties of the Reynell has not been studied in an autistic population, nor has the Achenbach Child Behaviour Checklist been validated specifically with PDD-NOS children. Although not relevant to a discussion on the effectiveness of treatment, the study’s researchers questioned the validity of the Family Satisfaction Questionnaire, as its psychometric properties were also untested in a population of developmentally disabled children.

**Summary of critical appraisal of Smith et al., 2000.** The stated purpose of this study was to extend the literature on the effectiveness of early intervention with PDD-NOS and to compare Lovaas regimens administered by either a trained therapist or by parents (both settings were in the home). Therefore, again there was an absence of a non-Lovaas control group in order to answer an effectiveness question. The study meets three of the six criteria (#2, 3, 6) with partial satisfaction of criteria #4 and 5. Therefore, methodologically Smith is a fairly strong study designed to ask a specific question.

This chapter has presented the findings from a health technology assessment of the legal and effectiveness domains of a judicial policy making case study (Auton). The next chapter will discuss the relevance of these findings by answering the research questions for this thesis.
CHAPTER 5: DISCUSSION AND CONCLUSIONS

Discussion

The Role of Evidence in Legal, Health Policy, and Health Care Contexts:
Seeking, Understanding and Applying Evidence

As can be concluded from the previous chapter, law, health policy and health care each have their own unique and divergent perspectives and processes for incorporating evidence into their professional lives. This research began with my initial conceptualization of these processes (Figure 2.1 & Appendix B) extrapolated from the 2000 IOM/AHRQ workshop: "Evidence": Its Meanings and Uses in Law, Medicine, and Health Care (Eisenberg, 2001). These concepts were framed within the legal context of a health technology assessment framework in order to examine the dynamics of a judicial policy making case study (Auton et al). After applying this framework to the data obtained from participant interviews and legal documents, the concepts were further refined to reflect these actors' conceptualizations of the identified themes. The findings from Chapter 4 will now be discussed with the focus of determining their relevance to the research question: "How do the sectors of law, health policy and health care seek, understand, and apply evidence?"

Perspective of Evidence & Focus of Application of Evidence

Perspective of evidence, the first contextual category of the evidentiary process, is the type of evidence utilized by each sector for decision making (e.g., population-based findings vs. individual-based, context-specific evidence) (Figure 4.1). Perspective of evidence also depicts the manner in which each domain conceptualizes the patient/client (Eisenberg, 2001). Focus of application of evidence is a related concept and pertains to the target of the outcome of the evidentiary process. Because of their linkage, these two categories will be discussed together.

Not surprisingly, the data from this research verifies law's individualist perspective of evidence. Judges decide cases based strictly on the merits of the specific evidence admitted to and weighed by the Court for a particular litigant; a judge's knowledge of other relevant evidence would not factor into the decision. Courts work within a "defined population of facts" which means that not all relevant evidence necessarily gets admitted; lawyers seek "persuasive" evidence to bolster an individual client's cause. This is done in isolation of the potential impact on society or policy. In the adversarial system, the judge has the final decision, but it does not necessarily mean that it is the right scientific decision. Judgments are the right decision, based on the evidence presented, for those particular parties, on one specific case, at that point in time. On appeal to a higher Court, a ruling is based on the state of evidence at the time of the original trial, despite any advances in science and changes in opinion that may have occurred since that time.
In regard to the focus of application of evidence, the legal data extends this category beyond its original individualist perspective with a couple of findings that occur as consequences of litigation. The first example of law adopting a population perspective occurs when a case is certified as a class action suit. This is an intentional strategy to apply a remedy received by a litigant to a specific class of the population. A second example is unintentional, and can occur when a judge accepts a specific argument as fact (e.g., that Lovaas Autism Treatment is effective), which results in a legal fact becoming a public endorsement that drives policy.

Within health policy, their perspective of evidence and focus of application remains population centred. The data did not reveal any examples of application of evidence at an individual level. In fact, the policy experts made reference to the importance of carefully examining stakeholder claims to ensure that the needs of the population are not compromised.

And pertaining to the health care domain, the data did not unveil any additional properties or dimensions of this category beyond what has already been reported.

Collectively, the concept of perspective of evidence remained stable across the three sectors; the data confirmed their original inherent perspectives of the patient/client. What has been expanded though is the category focus of application of evidence for the legal domain. Here the data showed that under certain conditions, legal actions do have an impact at the population level.

**Rules of Evidence**

Rules of evidence is the second contextual category that identifies the systems within each sector that are used to determine the acceptability of evidence. More specifically, rules of evidence are, "the ways in which evidence is brought to bear on the question at hand" (Havighurst et al., 2001, p. 199). Legal rules of evidence are explicitly articulated and codified; in addition, certain cases set precedence for acceptability of evidence. The claim that judges are granted great latitude in deciding what evidence to admit is supported by the data from this study. Since judges cannot be experts in all the substantive fields that approach the bench, the data would suggest that admitting evidence might be a less restrictive process than the weighing of evidence. Similar to the concept of burden of proof, it might be viewed as a "better" decision to admit five experts with questionable qualifications than to restrict one of exceptional standing. If so, a hypothesis might be that judges would have to wade through some junk science in order to get to the empirical evidence. This would require the honing of skills in appraising scientific evidence.

The legal rules of evidence are also in continual flux. This issue surfaced in the data with the example of the Courts redefining what constitutes standard of care (a "recognized school of thought", as opposed to what the "average" [e.g., at least 50% of the professional cohort] would do). One inherent risk with the various rules is their potential to restrict evidence (e.g., if
notice of intent to admit evidence is not submitted within specified time periods). The data identified factors such as a lawyer's experience, along with a client's financial ability to hire an expert to navigate the evidentiary maze, as ultimately impacting the evidence that makes it into Court.

The US workshop had not specifically addressed rules of evidence within the health policy context. Therefore, the data I discerned from this study provides the initial conceptualization of this category. One particular passage provided the basic concepts along with some of their properties and dimensions. Although this passage has been previously quoted in Chapter 4, its importance to the development of this category makes it worthy of being repeated:

There's an art in the science of being a good policy person . . . you have to be able to read the culture that you are in . . . in some governments . . . carefully collected, scientifically rigorous public opinion would be viewed as unimportant - because the public, you know, didn't have anything useful to say. So in that case, I would never put forward that kind of information as evidence for any particular opinion or position. So in certain cultures, political cultures, you come to see, to learn, what evidence is paid attention to and put that forward. (Participant: HP2)

The analysis of this paragraph identified "the art of being a policy person" (nested in the science of policy making) (a concept), with the ability to "read the [political and bureaucratic] culture" (a property or characteristic), in order to "know the type of evidence" (another property) that would be acceptable ("qualifying evidence" – another property) to the political party in power (the context), based on their value of each type of evidence.

In addition, another important property of a policy professional is "the ability to adapt and change" dependent upon the political party (a conditional dimension of the concept). While still conceptually undeveloped (and un-validated), this quote foretells an unstructured and conditional pattern to their rules of evidence.

In regard to the health care sector, the data validates a lack of agreement on formalized rules of evidence utilized by health care practitioners. In this sample, psychologists relied on their own professional experience, training, and mental "templates" to review research. However, it should be noted that the lack of reference to any specific appraisal tool does not imply these are not utilized in practice. Instead, the data indicates that, in this subgroup of health care practitioners, they initially rely on tradition and experience to appraise evidence. This evidentiary process is likely to differ between health care practitioners. For example, different outcomes would have been expected had a sample of medical doctors been interviewed.
And collectively, the analysis of this category reveals that rules of evidence is a dimensional concept, placing it along a continuum from highly formalized rules (law) to unstructured processes (health policy), with health care being somewhere in the middle.

**Process of Change**

Process of change is the third contextual variable of my framework and refers to the ways in which change occurs in each of the domains, along with how it impacts each conceptualization of evidence. In law, change occurs in response to a major decision by a judge or through legislation. In health care, scientific breakthroughs lead to early adoption by experts, and then later, diffusion to other clinicians. The IOM/AHRQ workshop did not describe the process of change in health policy.

As previously described, the data analysis expanded this category with the identification of five subcategories: (a) level of evidence required to implement change, (b) strategies for change, (c) reasons for change, (d) barriers to change, and (e) facilitators of change.

**Level of evidence required to implement change.** As this subcategory was only identified in the data of one sector (health care), it still requires further validation to determine its relevance to the main category. However, I feel it would be an important category to retain and investigate further as it could provide interesting data on concepts such as the “evidentiary threshold” for change. Medical doctors would be an ideal theoretical sample to initially explore since they are well versed in research; most likely this would yield frequent and easily recognizable key concepts.

**Strategies for change.** The data touched briefly on the strategies that each sector utilize to implement change. Again, law was the most formalized with their appeal process, while the health care data showed a more eclectic, less structured approach to change. The health policy data spoke of getting an issue onto the public agenda.

**Reasons for change.** The health care and health policy data identified some of the factors prompting change within each sector. These include political, occupational and professional forces. Why people (or professions) change is always an interesting concept. Why and how they change in relation to relevant evidence can have implications for plans for research dissemination.

**Barriers to change.** References to this subcategory were discovered only in the health care and health policy data. Health policy’s consensus that economic factors prevail as the greatest barrier was evident throughout the text. However, the health care data were void of economic references, instead focusing on occupational and personal reasons that prevent change. Each perspective is obviously reflective of the challenges faced in their daily practice.

**Facilitators of change.** An interesting finding in the legal data was the reference to the fact that at times, it is not the successfully litigated cases that affect change, but the
unsuccessfully litigated cases. One potential question would be whether this is analogous to the significance of null findings in scientific research.

**Adjudicators of Evidence**

Adjudicators, the "evidentiary actors" of my framework, are the experts who appraise and validate the evidence for their profession. The data analysis resulted in the identification of five subcategories, (a) experts used to provide opinion, (b) experts used to discredit opinion, (c) experts used to conduct research, (d) experts used for peer review, and (e) Court-appointed experts.

**Experts used to provide opinion.** Naturally, this category was heavily weighted towards the legal domain. The legal transcripts were rich with details on how lawyers pre-screen, select, instruct and interact with experts. Criteria for selecting expert witnesses not only include impressive curriculum vitae, but experts must also possess good communication skills in order to present a reasoned, articulate argument, with the ability to defend it on cross-examination. The data proved interesting in regard to the process of pre-screening experts. For example: (a) lists of "go to" witnesses, (b) reviewing experts' publication lists, (c) checking with colleagues, (d) ascertaining expertise (either training or experience), and (e) cost. Experts that are "hands on" in their field (as opposed to possessing only textbook knowledge), are more highly regarded, as are medical doctors with their objective medical evidence (psychiatry excluded).

**Experts used to discredit opinion.** Another interesting finding in regard to lawyers' cross-examination techniques was the distinction made between attacking an expert's factual matrix as opposed to attempting to discredit their opinion (an example of objective evidence vs. subjective evidence).

**Experts used to conduct research.** In Auton (2000) Allan J. concluded that the Lovaas effectiveness report commissioned by the Crown "exhibit[ed] an obvious bias towards supporting the Crown's position in the litigation" (para. 48). However, in the literature review for this thesis, critics of the adversarial system contend that both lay and expert witnesses called by each respective party would also naturally favor their attorney-employer's perspective. This scenario was posed during one of the legal interviews, along with the question as to how this context differed from lawyers securing and paying their own expert witnesses; the reply surmised that it probably was not much different.

**Experts used for peer review.** Although peer-review is a recurring concept in the health care data, no mention of it was found in either the legal or health policy data. The transcripts from the psychologists did not provide any additional information beyond what has already been covered in a preceding chapter.

**Court-appointed experts.** This concept was vehemently opposed by the legal professionals who argued that assigning experts to aid a judge in decision making went against the tenets of the adversarial system. In addition, there was a belief that there exists a potential to
introduce bias, since the Court expert would have direct access to the judge. According to the legal participants, Court-appointed experts are a very rare occurrence in BC Courts.

**Decision-makers of Evidence**

The final "evidentiary actor" of my framework, the decision-maker, has the ultimate responsibility of deciding what evidence their profession should follow. There was sufficient data to be able to discern five subcategories of this concept: (a) key players, (b) types of evidence, (c) level of evidence, (d) barriers to evidence, and (e) evidentiary basis for decision making.

**Key players.** The IOM/AHRQ workshop identified judges and juries as the key legal decision-makers; this was strongly supported in my data. The workshop also identified the criterion of credentials as an essential qualifier (a condition) for accepting an expert's opinion in the legal forum. The observation they made was that legal evidence is not judged in its own right; its validity is dependent upon the credentials of the expert. Their alternate argument was that in health care, evidence must be able to stand alone irrespective of who produces it; credentialing is not a relevant factor in assessing its validity. This is confirmatory with the legal experts interviewed for this thesis. Responses to interview questions that centred around the role of expert opinion typically contained references to "impressive curricula vitae", "qualifications", and "expertise". However, in the health care data, when questioned about levels of evidence required to change practice, typical responses focused more on the essential features of a study's design, conduct, and analysis. However, one respondent identified the study centre as an important variable in determining the validity of evidence. Therefore, a research centre's reputation could be a higher-order condition above the credentials of the researcher.

**Types of evidence.** Based on the types of evidence referred to by the participants of this study, law appears to have the broadest range to wade through: "junk science", affidavits, hearsay evidence, and precedent cases, right up the hierarchical ladder to randomized controlled trials. This highlights the need for legal professionals to be trained in research methodologies and quality assessment techniques. An interesting investigation would be to explore whether hearsay evidence is conceptually analogous to the medical custom of passing on knowledge through tradition (e.g., a medical student accepting the effectiveness of a procedure based on the "evidence" presented by their physician mentor).

**Level of evidence.** Another opinion to come out of the US workshop was that, although better educated, physicians typically seek synthesized information or secondary sources for their evidentiary needs, while the generally less-educated juries must rely on their limited skills to interpret highly technical research presented by expert witnesses. This specific assertion was difficult to verify from my data, as the health care professionals interviewed for this study were not medical doctors (hence, they may draw upon different levels of evidence). In addition, the context of the Auton case was a trial by judge, not jury.
Barriers to decision making. This subcategory emerged from the data as both the legal and health policy sectors voiced their frustrations in regard to factors that impede decision making. One lawyer commented that the type, quality, and amount of evidence they present to the Court impacts the judge's decision making ability. Under the rules of evidence, lawyers can be selective in what evidence they choose to reveal. This approach is contrary to a scientific method such as systematic review where the goal is a comprehensive search for evidence.

At a very basic level, disclosing information deemed beneficial to a case, while intentionally withholding potentially damaging evidence in the course of private litigation runs parallel to a much larger issue currently being debated - sequestered science (Givelber & Robbins, 2006). This phenomenon involves current legal techniques that are used to prevent the disclosure of potentially harmful information discovered at trial, which could be used in a public health sense to prevent disease, injury, disability, and death. The recent example is the suppression of evidence by Merck pharmaceuticals for their prescription drug Vioxx™. This phenomenon is also an example of law taking the individualist perspective.

Evidentiary basis for decision making. This subcategory, discovered in the health policy data only, acknowledges that evidence is only one small piece of the puzzle in regard to political policy making, and that the lack of evidence does not prevent policy from being developed. When a bureaucrat was posed the question, "When could politics trump evidence", the response was "always", since the politician was identified as the elected decision maker. One exception, however, was in regard to the Pharmacare process whereby the government adopted a strict evidence-based approach to drug approval. In this context, if there was not enough evidence to support an application, the drug would not be listed. Often times this occurred despite pressure from stakeholders and industry (this was an example of "evidence trumping politics"). What makes the drug approval process more quantitative and systematic is the fact that industry provides the required evidence (in the form of randomized controlled trials). In other health and social policy contexts, decisions would require the government to design their own research, which in most cases is not an option. However, no matter how good the evidence, a policy will not make it into practice unless there is public support, political support and the necessary financial resources.

Timing of Utilization of Evidence

This final category positions the patient/client outcome relative to the scientific evidence pathway (Appendix L). Graphically locating the "outcome" variable (represented by a circle with a cross through it) along each domain's pathway either before or after the evidentiary process, demonstrates each sector's theoretical position on utilizing evidence either before or after the outcome of interest. The data analysis did not reveal any additional information that would change the respective positions of this variable.
Application of Concepts to a Judicial Policy Making Case Study: Auton et al.

The concepts from my Scientific Evidence Pathway framework (Appendix L) were applied to the Auton trial in order to gain a greater perspective of how the evidence on Lovaas Autism Treatment was interpreted and applied at the various levels of the Courts. Although not all concepts were identified in the various Reasons for Judgment, there were a few indicators of some key points.

**Perspective of Evidence**

At the trial Court level, Allan J. accepted individual-based, context-specific evidence in the form of parental affidavits as proof of the “significant gains” made by the four child litigants. Yet, she discounted population-level data on Lovaas Autism Treatment submitted by the Crown (the BCOHTA report) stating: “it exhibit[ed] an obvious bias” (para. 48); along with rejecting a cost-benefit analysis of LAT presented by the petitioners declaring that “it [was] not possible to estimate accurately either the additional immediate costs of a treatment programme or the inevitable savings in the long run” (para. 145). In the end, Allan J. determined that the Lovaas evidence was not relevant, as the Court was not in the position to order one type of treatment over another. In effect, the Court’s decision to order EIBI as opposed to LAT was acknowledging a potential benefit for a population of autistic children as opposed to benefiting the needs of the specific children before the bench. However, at the Court of Appeal, although the decision held that EIBI as the program to be funded, it took on an individualist perspective by recognizing the effectiveness of LAT for the four children, and therefore modified the order so that they could receive the therapy. And finally, at the level of the Supreme Court of Canada, the perspective of evidence changed. No longer was the focus on the effectiveness of the therapy for the individual children. Instead the spotlight turned to a population perspective in regard to the specifics of a Charter challenge (discrimination against a population of autistic children).

**Rules of Evidence**

Although Allan J. provided a cursory critique of the BCOHTA report to substantiate her reasons for rejecting this evidence, no reasons were offered as to why the cost-benefit analysis was not considered credible. Justice Allan also applied the “general acceptance” rule of evidence in regard to the finding that other Provinces had adopted ABA treatment programs. As previously mentioned; policy precedence was now a surrogate marker for the purported effectiveness (or cost-effectiveness) of LAT.

**Process of Change**

Change was demonstrated throughout all three levels of the judicial court system beginning with the initial ruling for EIBI from the trial court judge. Next, at the level of the Court of Appeal, the judges’ assessment of the same evidence, along with the application of a new section of the Charter (s. 24) and an evaluation of prior jurisprudence, resulted in a change to the trial judge’s order to allow LAT for the petitioners. Then, at the Supreme Court of Canada,
the judges reverted back to an application of the language of sections 7 & 15, which overturned the lower Courts’ rulings. This scenario also contains indicators for the following subcategories for the process of change: (a) level of evidence required to implement change, (b) strategies for change, and (c) reasons for change.

**Adjudicators of Evidence**

The general role of experts in regard to providing opinion and discrediting evidence was previously described in Chapter 4. However, specific to Auton at the Supreme Court of Canada, was the addition of a different type of “expert” – interveners. An intervener is a person who is not an original party to the suit but someone who has a vested interest in its outcome. In the case of Auton, interveners included Attorney Generals of various provinces, in addition to representatives from community stakeholder groups and a person with autism.

**Decision-Makers of Evidence**

An interesting process occurred in relation to how the respective Court judges decided on the stage of diffusion of ABA/IBI/EIBI and Lovaas Autism Treatment. At the trial court level, the Crown presented the argument that LAT was still “an experimental therapy” (Auton, et al., 2000, para. 30). However, Allan J. did not acknowledge this in her analysis. At the Court of Appeal, reference was then made to the fact that the generic form of therapy (EIBI) (as ordered by the lower Court) was no longer viewed as experimental (this went undisputed). Then, at the Supreme Court of Canada, there were frequent references by McLaughlin, J. to the “novel”, “controversial” and “emergent” nature of ABA/IBI therapy (Auton et al., paras. 11, 56, 59, 61, 62), which in the Court’s opinion had “only recently began to be recognized as medically required”. This perspective apparently came as a surprise to the legal profession as it appeared this issue had been addressed in the lower Courts (personal communication, Participant L2, May 31/06). But then again, significant time had elapsed between the original trial in 2000 and the Supreme Court appeal in 2004. During this time period progress had been made and more provinces were integrating ABA into their autism programs. However, a point to recall is that the Court of Appeal reviews cases based on the evidence presented at the time of the original trial: “At the time of the trial in 2000, ABA/IBI funding for autistic children was only beginning to be recognized as desirable and was far from universal” (para. 11).

Because of the Supreme Court’s view that ABA/IBI was an emerging therapy, it rendered the comparator group that the petitioners had selected for their Charter claim invalid. Instead of comparing the child litigants to non-disabled children or adults with mental illness that receive funding for medically required treatments, the Court decided that the appropriate comparator would be a “non-disabled person suffering a disability other than a mental disability (here autism) seeking or receiving funding for a non-core therapy important for his or her present and future health, which is emergent [italics added] and only recently becoming recognized as medically required” (para. 55). Therefore, the Court’s responsibility was to apply the facts to an
appropriate comparator group. By doing this, a s. 15 challenge did not stand up. In this case, the Court's determination of ABA/IBI's level of diffusion had significant impact on the outcome of a Charter challenge.
Effectiveness of Lovaas Autism Treatment

The second half of this thesis was framed within the effectiveness domain of the comprehensive health technology assessment framework (Kazanjian, 2004) in order to answer the question, "What is the current state of knowledge in regard to the effectiveness claims of Lovaas therapy?"

Descriptive Data Synthesis

To determine the current state of knowledge on Lovaas effectiveness, the final stage of this systematic review required a descriptive data synthesis of the critically appraised studies. This synthesis focused on the between-group differences/variability (heterogeneity) across studies in regard to the following key elements, (a) participants, interventions and outcome measures (clinical heterogeneity), (b) study design and quality (methodological heterogeneity), and (c) reported effects (heterogeneity in results) (NHS, 2001). While some of the analysis criteria are generic (e.g., whether studies are randomized), others are more specific to the review question and population under study (e.g., age at diagnosis), and had been identified as important variables from previous research. The results of this data synthesis were then collated and are presented in tabular form in Appendix O as a qualitative analysis of between-group heterogeneity. When between-group differences were identified, hypotheses of direction of bias were proffered and are subsequently outlined in Appendix P.

Clinical Heterogeneity

Participants. Although the majority of the included studies reported no significant between-group differences in regard to sample size, sex ratio, diagnoses (when relevant), chronological age, IQ, and dropout rates, a few had identified significant differences in other potentially important pre-treatment variables such as parents' level of education and marital status. Cohen et al. (2006) noted that although these family variables have yet to be associated with outcomes, they might serve to influence parents' choice of treatment program in non-randomized studies. Perhaps well-educated parents research their options more thoroughly and request particular treatment programs. Therefore, as depicted in Appendix P, the direction of bias for the variable parental education was hypothesized to be in the direction of the experimental group for the two non-randomized studies (Cohen, 2006 & Howard, 2005) and was determined to be unknown for the one randomized study that reported this variable (Sallows & Graupner, 2005). In addition, the direction of bias for the criterion parental marital status is also listed as unknown at this point. Single parents may opt for programs that have minimal impact on their time commitment. Another example of between-group differences in Cohen is that the comparison group consisted of 40% PDD-NOS as opposed to 5% in the EIBT group. Since PDD-NOS is a milder diagnosis, these children may have had better outcomes than the AD group (Smith et al., 2000). However, the direction of the bias would in fact favour the comparison group, not the treatment group. And, in Howard et al. (2005) the non-intensive control group
comprised 44% PDD-NOS in contrast to 17% in the treatment group. However, the authors reported this as statistically non-significant (hence, why Appendix O shows the groups as being equal on this criterion). However, if the difference was significant, the direction of bias would again be towards the comparison groups. Howard’s treatment group was also significantly younger than both of their controls. Since research indicates that outcomes may be better if treatment is begun at a younger age, the direction of bias could be towards the treatment group. And finally, Sallows & Graupner (2005) maintained that their groups were similar at intake but based this claim solely on intake scores. Demographic information actually showed differences in the distribution of nonverbal children and parents’ level of education. However these differences were not tested for statistical significance. Some of the studies did implement statistical manoeuvres in an attempt to control for differences in pre-treatment variables, but Cohen et al., (2006) notes that, “statistical controls are not a satisfactory solution for pre-existing group differences, especially given the relatively small sample size” (p. S153). In the end, none of the studies can claim participant homogeneity.

When analyzing the heterogeneity across studies, the inclusion criteria were remarkably similar and included common variables such as, (a) diagnosis (AD and PDD-NOS), (b) IQ (range: 35-75), (c) age (range: 18 months to 7 years), (d) geographical proximity to clinic (when relevant), (e) and absence of concurrent medical conditions. Two studies also stipulated maximum hours of previous behavioural intervention prior to intake (Cohen et al., 2006 & Howard et al., 2005). In regard to diagnosis, the literature review showed a trend towards expanding intensive behavioural research to include the diagnosis of PDD-NOS. Pertaining to intake IQ, three of the studies set this criterion at >35. However, Eikeseth (2002 & in press) defined theirs to be >50, basing their decision on research that suggests children with intellectual functioning below this cut-off could be less likely to benefit from treatment. And, interestingly, Howard et al. (2005) did not include IQ in their eligibility criteria, yet reported no statistically significant differences between groups at intake. In the end, despite Eikeseth setting a higher IQ threshold, the mean intake IQs for all studies were similar and ranged from 50-61 for the treatment groups and 50-65 for the comparison groups. Also, it is noted that the criterion of geographical proximity to the research site could have produced selection biases.

The inclusion criterion of chronological age at intake was also similar across studies (range: 18-48 months), with the exception of Eikeseth et al. (2002 & in press) whose study’s purpose was to evaluate an older autistic population of 4-7 year olds. One variation that was discovered was in regard to intake age, where Cohen et al., (2006) used both the chronological age at diagnosis (18-42 months) coupled with the chronological age at intake (<48 months).

Interventions. Only two of the six studies incorporated formalized mechanisms within their design to measure treatment fidelity (Sallows & Graupner, 2005; Smith et al., 2000). While planned treatment durations were met across all studies, treatment intensity varied
considerably. Although both Cohen's and Howard's protocols called for 35-40 hrs/week (less for children under 3 years), the actual number of treatment hours received were not reported. Sallows & Graupner (2005), on the other hand, met their treatment intensity goal of 39 hours/week for their clinic-directed group; however, a surprising finding was that their parent-directed Lovaas treatment group received more therapy than was anticipated. In fact, the 32 hours/week resulted in a between-group difference of only 7 hours thus limiting their ability to look for treatment differences. This finding biased the results towards the parent-directed group. In addition, Smith’s treatment group only received 24 of its 30 hours/week thus producing a bias towards the comparison parent-training group.

A consistent finding across studies was the use of standardized Lovaas/UCLA-based treatment protocols. The only variation occurred with Howard et al. (2005) who followed Lovaas therapy guidelines as operationalized by Maurice et al. Also of interest was the initial use of aversive treatment on the first four children enrolled in Smith et al. (2000). This practice was subsequently terminated; none of the other studies utilized these methods. Therapist and supervisor training was also clearly outlined across all studies with one possible limitation noted in Howard et al. (2005) where their supervisors were only trained in generic ABA, as opposed to Lovaas therapy.

In regard to teaching formats employed by therapists, there was less consistency across studies. Four of the papers clearly described discrete trial training (DTT) as the predominant method. However, in addition to DTT, Sallows & Graupner (2005) also referred to the use of "procedures supported by subsequent research" (p. 422) while Howard et al. (2005) purported to use "general case programming and most-to-least prompt and prompt-fading" (p. 7). It is possible that these methods could have produced confounding effects. Another possible confounder in Sallows could have been the additional non-behavioural treatments received by 22 of the 23 children. This included modalities such as special education, speech therapy and sensory integration. And finally, pertaining to use of un-validated assessment tools (e.g., Reynell), since this particular tool was used consistently across all studies, any measurement bias would have been equally distributed.

**Methodological Heterogeneity**

**Study Design.** This small sample of studies does not contribute to resolving the methodological issue of lack of randomization. Although two of the six papers were randomized and controlled (Sallows & Graupner, 2005; Smith et al., 2000), randomization was not successful in one (Sallows & Graupner, 2005), and neither study utilized non-treatment control groups. Research has shown that non-randomized studies produce larger estimates of treatment effects than studies that utilize random allocation (Chalmers, et al., 1977; Schulz, et al., 1995). Similarly, if the allocation scheme is not properly concealed, though biases can occur in either direction, typically there is an exaggeration of treatment effects (Chalmers, et al., 1984). Hence, for this
review the four non-randomized studies along with the one RCT were assigned biases in the direction of their experimental groups.

Sallows & Graupner's (2005) study adhered to a treatment protocol approved by the National Institute of Mental Health. As previously described, the children were matched on pre-treatment IQ and then randomly assigned to either a clinic-directed or parent-directed Lovaas therapy intervention. Unfortunately, randomization was not successful. Also, since this study's purpose was to compare two intensities of Lovaas therapy delivered in two different settings, it did not incorporate a true control group into its design. The design also suffered from other methodological weaknesses. Both the Early Learning Measure and the Achenbach & Teacher form, along with all assessments of the moderate learners subgroup, were not blinded. Since there was no true control group, the direction of bias is difficult to discern.

Smith et al. (2000) was successful in implementing a randomized study. Here subjects were divided by diagnosis and then matched based on IQ prior to randomization; treatment allocation was concealed. However, again, there was no non-Lovaas comparison group. Instead, children were randomized to receive Lovaas therapy either from a therapist or from their parents (both settings were in the home). Smith also suffered from some of the similar methodological challenges as Sallows such as un-blinded assessment on some of the outcome measures. In addition, the treatment goal of 30 hours/week was not achieved due to staffing issues.

Another methodological feature that is important in behavioural research is the timing of the outcome measures so as to rule out educational and/or maturational progress. This was successfully controlled for as all the included studies had similar outcome time points.

Unfortunately, most of these studies (with the exception of Howard et al., 2005) continued to have small sample sizes. Hence, the resultant low statistical power to detect differences between groups was a possible reason for non-significant differences (or smaller differences than expected). Another limitation identified by Smith (2000) was the heavily tailed, skewed distributions that prevented some analyses.

Study Quality. Study quality was covered in detail in a previous chapter. However, to conclude from that analysis, each study's initial quality scores were translated into a qualitative descriptor of either 'low', 'moderate', or 'high' quality. It is important not to confuse quality of evidence with magnitude of effect, "the magnitude of observed benefits and/or harms from a service, although of critical importance to decisions about whether it should be recommended, is a separate issue from the quality of the data" (Harris, et al., 2001, p. M-15).

Due to methodological limitations highlighted by the application of the Reisch and Jadad quality assessment tools, four of the six studies were determined to be of poor quality (Cohen et al., 2006; Eikeseth et al., 2002 & in press; Howard et al., 2005), while the two RCTs (Sallows & Graupner, 2005 & Smith et al., 2000) were assessed as being of moderate quality.
overall. Although Sallows & Graupner (2006) and Smith et al. (2000) initially achieved moderate and high Jadad scores (4/7 and 6/7 respectively), these were somewhat misleading as their quality and relevancy decreased once more comprehensive critical appraisal criteria were applied. A discovered limitation of the Jadad tool was that by not being able to assign partial marks, Smith et al.'s score became artificially inflated. Similarly, a strict reliance on study design hierarchy alone leads to an overestimation of a study's actual quality. Therefore, it is imperative that each study also be assessed based on specific design criteria.

Heterogeneity in Results

**Reported Effects.** As Appendix O depicts, the most homogenous outcome across studies was intellectual testing, with five of the six reports claiming statistically significant gains for their treatment groups. Sallows & Graupner (2005) were the only exception since, as previously described, the parent-directed group received more therapy hours than expected, thus making it difficult to detect any between-groups treatment effects. However, a post hoc subgroup analysis (combining the parent-directed group with the clinic-directed group and comparing their pre and post IQ scores) uncovered a statistical within-group difference. Another subgroup analysis compared the rapid learners with the moderate learners and noted significant gains in IQ for the rapid learners group.

Other statistically significant outcomes across studies included the Merrill-Palmer (visual-spatial) (two studies); Reynell (total) (language) (3 studies); and the Vineland composite (adaptive behaviour) (4 studies). Other measures such as the Vineland Maladaptive, Achenbach (Teacher Report Form) (behaviour), Learning Rates, Woodcock-Johnson (academic) and the Early Learning Measure (rates of learning) were only used by individual studies; therefore, aggregate results could not be reported. In addition, only two studies reported data on regular classroom placement rates. Cohen et al. (2006) stated that 17 of their 21 EIBT children (81%) were integrated into regular classrooms. Of these, six (28%) were fully integrated without assistance, four (19%) were fading out their shadow tutor, and seven (33%) required a full shadow. In addition, one child in the comparison group (5%) was placed in a regular classroom (their level of support was not reported). Smith et al. (2000) reported that six students (40%) in the intensive treatment group progressed into a regular classroom, of which four (27%) did not require assistance, while two students (13%) had a tutor. In their parent training comparison group, none of the children were assimilated into a normal classroom without assistance and three of the children (23%) attended regular classes with support. Although only based on two studies, both Cohen and Smith show similar normal classroom assimilation rates (without support) of 28% and 27% respectively.

**The Current State of Knowledge on Lovaas Effectiveness Evidence**

Over twenty years has passed since Lovaas' model of applied behavioural analysis was first introduced. Since then, researchers have designed, refined, operationalized, and controlled
studies in search of the ideal intensity, provider, training, setting, instructional technique, duration, population, timing and outcome measures. While many autism therapies have come in and out of favour since then, there remains a seductive allure to Lovaas therapy – precipitated by the original 1987 study proclaiming a chance for some autistic children to achieve “normal functioning”. It is no wonder this therapy has retained a faithful following of supporters all these years. But, where are we now in regard to our current state of knowledge on Lovaas effectiveness evidence? Answering that question required a retrospective review of previous Lovaas therapy critiques, contrasted against the critical appraisal data discerned from this systematic review.

This next section will describe advances and improvements in the field of Lovaas research since the original 1987 study and the last systematic review (Bassett et al., 2000). Additionally, ongoing methodological issues are summarized and unanswered research questions or future research needs are identified.

Advances and improvements. This descriptive data synthesis serves to support the a priori assumption that, advances in Lovaas research have done little to advance the effectiveness debate and in fact, with the identification of new confounding variables, has probably resulted in an even more complex phenomenon to investigate. For example, researchers are now turning to questions that address the identification of potential prognostic indicators of success, in addition to whether residual symptoms of autism remain after achieving normal post-treatment test scores (Sallows & Graupner, 2005). Smith, Cohen, and Howard expanded their research studies to include children with PDD-NOS. Eikeseth applied Lovaas therapy to an older cohort of 4-7 year olds, while most of the studies sought to improve external validity by situating the therapy within more naturalistic settings such as the child’s home and school.

Improvements in Lovaas techniques include the elimination of the most controversial aspect, which was the use of aversive consequences. In addition, the studies generally include improved reporting on treatment protocols, staff/parent training, and quality control (when implemented). However, it was difficult to discern whether this progress constituted stronger study designs or merely an improvement in reporting standards. Outcome measures are also more standardized and comprehensive, and in some cases have obviously benefited from twenty years of refinement. None of the studies utilized the disputed pro-rated mental age measurement that Lovaas had originally adopted. In addition, most researchers now recognize the gender bias and control for this when possible. And, finally, whether it can be called an advancement or improvement may be open for debate, but it is noted that none of the studies refer to their best outcome children as “normal”; instead, opting for the description of a child “reaching average range”.

Ongoing methodological issues. Many of the methodological issues that plagued the original Lovaas study are still evident in the current research. Problems such as referral and
selection bias; un-blinded outcome assessors; and, lack of treatment fidelity monitoring, continue to threaten both the internal and external validity of the results. Randomized trials designed without non-Lovaas control groups may be reflective of general acceptance of Lovaas efficacy within certain research circles. Lack of true control groups may also be indicative of a push towards advancing more of a “real world” agenda by studying variables such as intensity (to address budget concerns), provider training, and treatment settings as opposed to broadening the effectiveness data. As previously described, a paucity of randomized trials has also been linked to legal and ethical issues in research, such as Court decisions, highly educated parents refusing randomization, and mandates of treatment/research centres (Cohen et al., 2006).

**Included studies’ reported limitations.** The study authors identified the following limitations of their own research:

**Cohen et al., 2006**
1. Non-random assignment
2. No measure of treatment fidelity
3. Emphasis on testing developmental level as opposed to features of Autistic Spectrum Disorder

**Eikeseth et al., 2002 & in press**
1. Small sample size
2. Quasi-random group assignment
3. More cognitive measures as opposed to social development measures
4. Lower treatment intensity
5. No direct quality control measures of treatment

**Howard et al., 2005**
1. Non-random assignment
2. Un-blinded outcome assessors
3. Outcome measures focused on formal, standardized, norm-referenced assessments instead of also incorporating “the repeated direct observational measurement of behaviour in-situ that characterizes applied behaviour analysis” (p. 15).
4. Between-groups statistical comparisons not revealing clinically significant changes for an individual child
5. Treatment integrity not measured

**Sallows & Graupner, 2005**
1. Different IQ measures pre and post treatment
2. Un-blinded assessor for moderate learners group
3. Unequal samples (despite matching and then random assignment)
4. Small sample size

Smith et al., 2000

1. Small sample size, with heavy tailed, skewed distributions
2. Only one measure of social skills (Vineland)
3. Parent satisfaction questionnaire with untested psychometric properties
4. No measure of child's or parent's quality of life
5. No measure of parent's participation in treatment
6. Instruments not designed specifically for children with developmental disabilities
7. Lack of a standardized diagnostic instrument (and no follow-up diagnostic assessment)

**Unanswered research questions.** The systematic review process is instrumental in making transparent what we know and what we do not know about the topic of interest. A review of the included studies determined that the following research questions remain unanswered in regard to Lovaas Autism Treatment:

1. What is the "best" intensity?
2. What is the "best" setting?
3. Whether some children become indistinguishable from typically developing peers.
4. Whether measures of children's skill acquisition early in treatment are predictors of outcome.
5. Whether IQ and language are predictors of outcomes.
6. Whether parent involvement improves outcomes.
7. Whether results can be replicated in other studies with 4-7 year-olds.
8. Whether younger children would benefit if the intervention were delivered at school rather than at home.
9. Whether specific intake variables are associated with individual differences in response to treatment.
10. Whether age at intake predicts treatment outcomes.
11. Whether child, family, and treatment variables are correlated with outcomes.
12. Whether "eclectic" intervention is effective.
13. Whether "advanced social skills training" is effective.
14. Whether other teaching methodologies besides discrete trial training are effective (e.g., video modeling, incidental teaching).
15. Whether models that predict if a child would be a rapid or moderate learner could be validated.
16. Whether there is a correlation between the acquisition of social skills and the amount and duration of supervised peer play.
Strength of Body of Evidence

The final step in the descriptive data synthesis was grading the strength of the body of evidence. Strength of evidence is comprised of three domains, (a) quality, (b) quantity, and (c) consistency (AHRQ, 2002). Within an evaluative context, quality is, “the aggregate of quality ratings for individual studies, predicated on the extent to which bias was minimized”; quantity is, “numbers of studies, sample size or power, and magnitude of effect”; while consistency is, “the extent to which similar findings are reported using similar and different study designs” (Lohr, 2004, p. 11). The included studies were appraised in respect to these three domains, in addition to an examination of internal validity (the ability of the studies to produce valid results pertaining to the populations and settings in which they were conducted); and external validity (whether the studies were relevant and generalizable to a larger population) (Lohr, 2004).

The grading system developed by the Cochrane Collaboration (Higgins & Green, 2006) was used to evaluate the strength of inference about the effectiveness of Lovaas therapy. This system was identified in two systematic reviews of strength of evidence tools (AHRQ, 2002; Lohr, 2004). These reviews determined that the Cochrane tool adequately addressed the requisite quality, quantity, and consistency domains. In addition, the tool had been specifically designed to promote evidence-based health care, as opposed to being used to develop guidelines or practice recommendations. Hence, the following criteria were examined:

1. The quality of the included trials
2. The size and significance of the observed effects
3. The consistency of the effects across trials
4. Dose-response relationships
5. Other plausible competing explanations for the observed effects

Quality of the included trials. Both conceptually and practically, quality and strength are related, albeit hierarchical, ideas. One must grade the quality of individual studies before one can draw affirmative conclusions about the strength of the aggregated evidence (Lohr, 2004, p. 12). Based on Reisch quality assessment scores, the four observational studies (Cohen, 2006; Eikeseth, 2002; Eikeseth, in press; and Howard, 2005) were determined to be of poor quality. Initially the two randomized studies (Sallows & Graupner, 2005; Smith et al., 2000) were judged to be of moderate and high quality respectively, however, as previously mentioned, Smith was downgraded to moderate later in the analysis due to identified methodological limitations. Pertaining to the observational studies, the mean objective Reisch score (“Rating 2”) across studies was only 0.37 (a ratio score). As an aggregate, the six included studies were determined to be of poor quality.

Size and significance of observed effects. The primary outcome measure for all studies was IQ, with five of the six studies reporting statistically significant increases for their experimental groups. Although Sallows & Graupner did report significant differences once groups were
combined and divided into rapid and moderate learners, there were no significant differences when initially analyzed as per their a priori group assignment. Post hoc analyses are inherently observational and as such are subject to bias through confounding by other trial characteristics; therefore, the results cannot be confidently interpreted.

Performing numerous post hoc subgroup analyses to explain heterogeneity is data dredging. Data dredging is condemned because it is usually possible to find an apparent, but false, explanation for heterogeneity by considering lots of different characteristics. (Alderson, 2004, p. 121)

Table 5.1 below outlines the mean between-group difference at follow-up.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ gain</td>
<td>14</td>
<td>10</td>
<td>15</td>
<td>24</td>
<td>NSS</td>
<td>16</td>
</tr>
<tr>
<td>Merrill-Palmer</td>
<td>NSS</td>
<td>NSS</td>
<td>NR</td>
<td>24</td>
<td>NSS</td>
<td>15</td>
</tr>
<tr>
<td>Reynell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Comprehension</td>
<td>NSS</td>
<td>11*</td>
<td>NR</td>
<td>22</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td>- Expressive</td>
<td>NSS</td>
<td>18*</td>
<td>NR</td>
<td>23</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td>- Total</td>
<td>NR</td>
<td>15*</td>
<td>NR</td>
<td>NR</td>
<td>NSS</td>
<td>26</td>
</tr>
<tr>
<td>Vineland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Communication</td>
<td>-16**</td>
<td>12</td>
<td>22</td>
<td>19</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td>- Daily Living</td>
<td>-10**</td>
<td>NSS</td>
<td>16</td>
<td>9</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td>- Socialization</td>
<td>NSS</td>
<td>NSS</td>
<td>14</td>
<td>9</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td>- Motor</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>9</td>
<td>NSS</td>
<td>NR</td>
</tr>
<tr>
<td>- Composite</td>
<td>13</td>
<td>7</td>
<td>18</td>
<td>12</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td>- Maladaptive</td>
<td>N/A</td>
<td>N/A</td>
<td>-4.7</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* Incorrectly reported in primary study
** Estimated by graphical data
NSS - Not statistically significant
NR - Not reported
N/A - Not applicable

As Table 5.1 depicts, the variable of IQ ranged from a gain of 10 points (Eikeseth, 2002) up to 24 points (Howard, et al., 2005), with an average mean gain across the five studies of 15.8 points. The next statistically significant outcome measure was the communication subscale of the Vineland with a mean between-group difference at follow-up ranging from 12 points (Eikeseth, 2002) to 22 points (Eikeseth, in press), averaging 17.2 points across the studies. The total Reynell ranged from 15 points (Eikeseth, 2002) to 26 points (Smith, 2000), with a mean of 20 points. However, as previously explained, there appeared to be some inaccuracies in Eikeseth's (2002) tables, placing their results in question. And, lastly, the Vineland composite mean between-
group difference at follow-up ranged from 7 points (Eikeseth, 2002) to 18 points (Eikeseth, in press), with an average 12.5-point gain across studies. All the other statistically significant outcomes occurred in isolated studies; therefore they were too few in number to compute averages.

**Consistency of effects across trials.** The epidemiological principle of consistency is, "the extent to which diverse approaches, such as different study designs or populations, for studying a relationship or link between a factor and an outcome will yield similar conclusions" (AHRQ, 2002, p. 23). Unfortunately, an accurate assessment of whether both similar and different study designs produced comparable outcomes could not be done due to the lack of quality research on this topic.

**Dose-response relationships.** Based on this limited data, there does not appear to be a dose-response relationship between intensity of treatment and follow-up IQ measures. Smith et al. (2000) reported an intensity of 24 hours per week of EIBT with a mean between-group difference of 16 points; while Cohen et al. (2006) described their experimental group as having received between 35 and 40 hours of therapy yet only achieving a mean between-group difference of 14 points. There was also no dose-response relationship in regard to the communication subscale of the Vineland measure. Here, Eikeseth (2002) initially reported a 12-point mean between-group difference after one year of treatment at an intensity of 28 hours per week. Yet, upon follow-up at age 8, when intensity had been reduced to 18 hours per week (since the age of 6), the mean between-group differences had risen to 22 points. And finally, when reviewing the Vineland composite scores, the same holds true with Eikeseth (2002) initially achieving the lowest mean between-group differences at a reported 7 points (at an intensity of 28 hours), then achieving the highest mean between-group difference at follow-up with 18 points (with an intensity of 18 hours).

**Other plausible competing explanations for the observed effects** This final criterion included ruling out chance, bias and confounding. Chance is assessed by an examination of the p-values and confidence intervals (when reported). P-value interpretation was limited because of its dependence on sample size. In this cohort of studies, all but one had an n<30. Two of the studies reported confidence intervals (Cohen, et al., 2006; Eikeseth, in press); however, the intervals were fairly wide, representing more opportunity for chance variation (a definitive study would have a significant p-value and a narrow confidence interval).

Bias is "any systematic error . . . that results in an incorrect estimate of the association between exposure and risk of disease" (Hennekens & Buring, 1987, p. 272). Conditions that support valid results would be evidence that the researchers attempted to control biases in design, conduct, and analysis. And finally, confounding is "the mixing of the effect of the exposure under study on the disease with that of a third factor" (p. 287). This factor is associated with the exposure and is a risk factor for the disease. One known confounder in autism is the sex
of the child (with girls having a poorer prognosis). However, since little is known about the etiology of autism, unknown confounders may have impacted study results. This makes it even more imperative for investigators to control for potential confounding through procedures such as matching, randomization, and the analytical technique of stratification.
Conclusions

A retrospective analysis of the legal context of a health policy decision was designed to answer two specific research questions: (a) How do the sectors of law, health policy and health care seek, understand and apply evidence?; and, (b) What is the current state of knowledge in regard to the effectiveness claims of Lovaas therapy? (Appendix A). Both of these objectives had been approached through the application of a health technology assessment framework, specifically the legal and effectiveness domains.

From a conceptual perspective, the first research question was posed in order to provide policymakers and legal professionals with a greater understanding of each other’s divergent conceptualizations of the evidentiary process. This difference became apparent within a local context during the Auton legal proceedings on the provision of provincially funded Lovaas Autism Treatment.

As Appendix L illustrates, each sector holds a specific perspective of evidence. As their adjudicators and decision-makers (experts) travel back and forth along their respective scientific evidence pathways, navigating their way through various abstractions of rules of evidence and processes of change, their perspectives may be challenged when their pathways cross with another sector’s. If the other sector’s perspective of evidence prevails, this could alter their focus of application of evidence (evidence applied to either an individual or to society in general). In this thesis, this process took place under the context of judicial policymaking.

On a practical level, what additional knowledge has been achieved through this study? What should legal and health policy actors know about each other’s understanding and use of evidence? Although health care and health care professionals were also included in this analysis, and the three-way analysis contributed to delineating similarities and differences, the data were heavily weighted towards the tension between law and policy. Therefore, this will be the focus of my conclusions.

Law should know that in health policy there is no “objective and dispassionate” method to decide policy. Policy is developed after consideration of a multitude of factors; evidence being just one small piece. In many cases, evidence does not exist for a particular policy problem; therefore alternate forms of information must be substituted. When evidence is available, its relevance is often dependent upon the particular politics of the governing party; with some parties valuing specific forms of evidence, such as public opinion, and others placing more credence elsewhere. However, quality of evidence is irrelevant if there is no perceived or voiced public support and political willingness to fund a program. In addition, as the elected decisionmakers, politicians have the ultimate authority to go against the evidence.

Health policymakers should know that in law, the search for evidence is not necessarily comprehensive. Unlike health care, where systematic reviews of synthesized data both for and against an intervention are sometimes used for decision-making, law seeks evidence in support
of their particular client's claim. Although a comprehensive search may be conducted, only the relevant evidence would be disclosed. This is not necessarily a criticism of the process, but merely an example of law's individualist perspective that protects the rights of their client (assumingly, in the spirit of the adversarial system, the opposing lawyer would disclose any contrary evidence).

A health policy-maker may also be interested in knowing that potential expert witnesses are thoroughly and systematically researched and screened prior to trial. By accessing specific legal databases, lawyers can determine whom experts have testified for, what they said, and whether it met with the judge's approval. As an expert witness, the objective is to persuade the judge to accept their opinion. However, rivalling this objective is an opposing lawyer who attempts to establish bias and/or find fault in the factual matrix of an expert's testimony (as opposed to challenging the expert's actual opinion). While there are suggestions that Court-appointed experts are an effective alternative, others point out the risk of bias established by their direct link to a judge.

And, in regard to the role of the judge, health policy-makers should know that a judge's ruling is dependant upon the quality of evidence before them. Although the judgment may not be construed as "right", in most cases it would be considered the right judgment based on the evidence at hand. Although it may be apparent retrospectively that evidence was missing or weak, judges are discouraged from becoming involved in seeking additional information by cross-examining witnesses. Also, another important finding was that a judicial endorsement of an issue might become a public fact; the example being, Allan J.'s acceptance of the effectiveness of Lovaas therapy for the specific child petitioners resulting in a perceived credibility to the treatment in the public eye.

The second research question sought to determine the current state of knowledge in regard to Lovaas Autism Treatment. A systematic review and critical appraisal of the evidence were the HTA methods employed to update a previous review (Bassett et al., 2000). A literature search that spanned the period of January 1998 to March 2006 initially recovered 285 citations, of which 58 were deemed relevant by preliminary criteria. However, once full inclusion criteria were systematically applied, only six studies remained. A quality assessment of these six remaining studies, utilizing design-specific assessment tools, highlighted varying methodological quality. The four observational reports achieved a mean aggregate quality score of just 0.37 (poor), while the two RCTs were determined to be of moderate quality.

The subsequent critical appraisal, which employed the criteria identified by the American Autism Working Group, confirmed that most of the studies still contained the same methodological limitations previously identified in the original 1987 Lovaas study. In fact, on average the studies only met 2.5 of the 6 criteria identified as essential components for future behavioural and social intervention research. And finally, a qualitative analysis of heterogeneity
across studies confirmed the presence of clinical heterogeneity of participants in that none of the groups were similar at intake (although often reported otherwise). Most likely this resulted in a net direction of bias in support of the experimental treatment groups. This heterogeneity threatens external validity in regard to the confidence in which results can be generalized beyond this study.

Additional heterogeneity across studies also existed in relation to the experimental intervention, as evidenced by poor monitoring of treatment fidelity, limited use of non-Lovaas control groups, and various standards for therapist training. Also of interest is that none of the studies were direct comparisons between Lovaas and other ABA interventions. However, when all these factors are considered, the net direction of bias appears to be neutral; yet, these limitations still serve to threaten the internal validity of the studies.

In regard to methodological heterogeneity across studies, the greatest identified weakness in the majority of the studies was the lack of randomization, along with un-blinded outcome assessment and treatment allocation. Again, this can result in a bias in favour of the experimental group. These are considered serious flaws in study design.

The most consistent statistically significant outcomes across the six studies were that, on average, Lovaas Autism Treatment increased children's IQ by 15.8 points, Vineland communication scores by 17.2 points, Vineland composite scores by 12.5 points, and Reynell Total scores by 20 points. However, methodological weaknesses, post hoc analyses and incorrect data reporting make it difficult to assert that this collective body of evidence advances Lovaas effectiveness claims; at best they contribute to hypothesis generation.

Overall, the strength of the body of evidence on Lovaas Autism Treatment was determined to be poor. Research papers of poor to moderate quality, coupled with few studies from the top of the research hierarchical ladder, and lack of consistency in outcomes weakens this small group of studies. While it is probable that increases did (and can) occur in the reported measures, the magnitude of effect, along with the intensity and setting required to achieve it, is still in question. At best, these studies show small effects in favour of Lovaas Autism Treatment. It would be up to policy-makers and clinicians to determine whether these outcomes are clinically significant enough to warrant the expense of implementing this specific form of autism treatment. Nevertheless, this critical appraisal has provided the requisite information for clinicians and policy-makers to be able to judge its applicability to their own unique circumstances.

Together, the results from the two analyses in this thesis have contributed to a greater understanding of how evidence was used by each respective party in the Autism et al. legal proceedings. For policy-makers, this research provides insight into the judicial policy making process and the importance of transparency in policymaking. In addition, the systematic review has updated the empirical evidence on the effectiveness of Lovaas Autism Treatment (which could be relevant for future policy considerations). For health technology assessment
researchers, this thesis highlights a qualitative approach (Grounded Theory) for the examination of the legal context of the HTA framework. This methodology serves to broaden the scope of evidence HTA researchers can offer. This research also features the incorporation of a qualitative data analysis software program into the HTA framework. Its utility could prove beneficial for the synthesis of qualitative research into HTA projects.

This thesis has conceptualized, defined, and then refined the phenomenon of the role of evidence in regard to its properties and dimensions. The next step is to conduct further theoretical sampling to explore hypotheses and statements of relationships between concepts.

**Strengths and Limitations**

One strength of this research was the comprehensive search strategy utilized. The selected search terms were effective in identifying only the essential studies by screening out the myriad of alternative study designs. The results of this search strategy were verified by one of the leading researchers in the field – Dr. Tristram Smith – thus providing a level of confidence that all the relevant research had been recovered. Another strength was in the techniques utilized to more objectively assess the qualitative aspects of the retrieved studies during the critical appraisal phase. This included tabulation of features of heterogeneity (Appendix O: Analysis of Heterogeneity Between Groups), and direction of biases (Appendix P: Direction of Biases). While the biggest strength of the research was the opportunity of taking preliminary conceptualizations of the role of evidence and being able to study them within the context of judicial policymaking.

One of the limitations of the study was in regard to not being able to interview medical practitioners. As previously mentioned, their perspectives on the role of evidence would have been more representative of the health care domain than those of the psychologists. Additionally, being able to interview a judge would have also rounded out the legal perspective of evidence. An obvious limitation of this study was the poor yield of methodologically sound research studies. And finally, the legal document: Reasons for Judgment was not found to be a very rich source of data. A future approach could be to expand the search to include other documents such as factums and affidavits, as these may be more illuminating about the processes used to argue the evidence.

**Further Research**

As described in the methodology chapter, grounded theory was not employed to its fullest extent since the emerging state of knowledge on the phenomenon of the role of evidence precluded the higher level of abstraction - theorizing. Instead, conceptual ordering was utilized to build on a preliminary hypothesis. The findings from this research, along with the conceptualization of the three distinct scientific evidence pathways, could be expanded through theoretical sampling to further develop categories along their properties and dimensions. From there, relating categories through hypotheses could be achieved.
One particularly interesting area for future research would be in regard to the rules of evidence for policymaking. As described under this title at the beginning of this chapter, one specific quote extracted from the interview data revealed some interesting concepts, actions and interactions between bureaucratic and political cultures.

The subcategory level of evidence under the category of process of change is another possible area for future research. Sampling medical physicians in regard to the type of evidence they seek for decision-making could possibly add a dimensional quality to this category – an "evidentiary threshold".

The two in vivo (verbatim) codes that were identified under the reasons for change subcategory are also intriguing. Firstly, law's reoccurring theme of "persuading" evidence, and health care's reference to being "influenced" to change.

And finally, an additional lens with which to view the role of evidence, and bring another perspective, could have been from the vantage point of the parents of these autistic children.

Post Script

On April 12, 2007, the Supreme Court of Canada dismissed the Wynberg and Deskin families' application for leave to appeal the Court of Appeal for Ontario's ruling (Wynberg, R. et al v. H.M.Q. in Right of Ontario and Deskin, M. et al v. H.M.Q. in Right of Ontario). Therefore, it was held that the government of Ontario was not discriminatory by failing to provide an Intensive Education Intervention Program to autistic children age six and over.
REFERENCES


Summers, R.S. (1999). Formal legal truth and substantive truth in judicial fact-finding: Their justified divergence in some particular cases. Law and Philosophy, 18, 497-511.


Williams, A. (1997). All cost effective treatments should be free . . . or, how Archie Cochrane changed my life! Journal of Epidemiology and Community Health, 51, 116-120.


APPENDICES

Appendix A
The Role of Evidence in Legal, Health Policy and Health Care Contexts

Evidence

Input

Domains

Law

Health Policy

Health Care

Seeking, Understanding, and Applying Evidence
(See Appendix L: Scientific Evidence Pathways)

Legal Review of Evidence

Policy Review of Evidence
Case study: Lovaas Autism Treatment

Medical Review of Evidence

Judicial Policy Making

Role of Evidence Effectiveness of Lovaas

Health Policy
Appendix B
Scientific Evidence Pathways: Initial Framework

Perspective of Evidence
- Individual (Client)
- Population (Society)
- Population (Society)

Level of Evidence
- Primary Evidence
- Secondary Evidence
- Secondary Evidence

Rules of Evidence
- Codes & Precedence
- Critical Appraisal
- ?

Gatekeepers of Evidence
- Lawyers
- Judges
- Journal Editors
- Bureaucrats

Decision-Makers of Evidence
- Judges
- Key Opinion Leaders
- Bureaucrats & Politicians
- Policy Debate

Change Agents of Evidence
- Legislation
- Key Opinion Leaders
- Population (Health Care System)

Level of Application of Evidence
- Individual (Client)
- Individual (Patient)
- Population (Health Care System)

GENERALIZABILITY OF OUTCOMES TO THE POPULATION (SOCIETY)

Conceptualized from:
Eisenberg, 2001
Appendix C
Participant Letter of Introduction

The University of British Columbia
Department of Health Care and
Epidemiology
Faculty of Medicine

Date

Dear

My name is Sandra Beard and I am a doctoral student in the department of Health Care and
Epidemiology in the Faculty of Medicine at UBC. I am under the supervision of a thesis
committee headed by Dr. Arminee Kazanjian.

Part of my thesis involves an examination of the role of scientific evidence within the contexts of
law, policy and health care and its impact on health policy. In particular, a component of this
research is a case study of the Auton et al. legal proceedings in regard to public funding for
Lovaas autism treatment for children.

As you have been identified as a key opinion leader with expertise and/or familiarity with this
case, I would like to request your participation in this research by way of an interview. In addition
to supporting doctoral research, this would provide you with the opportunity to provide your
own perspective on a controversial topic. Of course, anonymity will be assured.

Enclosed you will find a participant Consent Form. It is not necessary to sign it at this point; it is
merely provided for your information.

I would appreciate it if you could contact me at your earliest opportunity to inform me of your
decision either way. I can be reached at (phone #). Alternatively, my email address is: (email
address)

Sincerely

Sandra Beard, RN, MSN, PhD(C)
UBC HCEP Doctoral student
Appendix D
Consent Form

The University of British Columbia
Department of Health Care and Epidemiology
Faculty of Medicine

The role of evidence in legal, policy and health care contexts: Autism and the impact of legal decisions on public funding for autistic children.

Principal Investigator: Dr. Arminee Kazanjian
Department of Health Care and Epidemiology
Faculty of Medicine, UBC
O: (phone #).

Co-Investigator: Sandra Beard, RN, MSN, PhD(C)
Doctoral Student
Department of Health Care and Epidemiology
Faculty of Medicine, UBC
O: (phone #)

The following research study is a partial requirement for the completion of Sandra Beard’s doctoral thesis.

Purpose:
You have been identified as a participant in a study investigating the role of scientific evidence in legal, policy and health care contexts. Specifically, one of the purposes of this thesis is to examine the differing conceptualizations of evidence in a legal Court case on Lovaas treatment (Auton et al.) and the impact on health policy. You have been asked to participate in this study because of your expertise and familiarity with this case.

Study Procedures:
If you decide to participate, you will be interviewed one-on-one by the Co-Investigator on the topic of the role of evidence unique to your professional philosophy. The interview can take place in your office or another location that is most convenient. The process will entail responding to open-ended questions, of which you will have the ability to refuse to answer if you so desire. For analysis purposes it would be preferable if the interview is audio-taped, however, hand-written notes could be taken if you prefer. The research will constitute a single interview, lasting approximately one hour, with the possibility of a follow-up interview/telephone call if further clarification is required.

Confidentiality:
Audiotapes or research notes will be coded to protect your identity. They will be stored in a locked file cabinet in the Co-Investigator’s office. Transcriptions of audiotapes will only be shared with the Co-Investigator’s thesis committee and research assistants. Results of the study will be published. However, identifying features will not be mentioned. If a quote (verbatim or paraphrased) is used, you will only be identified in a general manner that could not be linked directly to you. You will appreciate that, as the holder of an important position in the Public Sector, some risk of indirect attribution may remain.
Contact for information about the study:
If you have any questions or desire further information with respect to this study, you may contact Dr. Arminee Kazanjian at (phone #).

Contact for concerns about the rights of research subjects:
If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604-822-8598.

Consent:
Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without jeopardy to your present or future relationship with the University of British Columbia.

Your signature below indicates that you have received a copy of this consent form for your own records.

Your signature indicates that you consent to participate in this study.

I agree/do not agree to have the interview audio-taped (please circle answer)

Subject Signature Date

Printed Name of the Subject signing above.

Co-Investigator Signature Date

Printed Name of Co-Investigator signing above.
Appendix E
Search Strategy

Electronic Databases
Medline (OVID)
1998 to November Week 3 2004
Database: Ovid MEDLINE(R) <1996 to November Week 3 2004>
Search Strategy:

1. exp Autistic Disorder/ (3203)
2. (autis: or asperger: or kanner: or lovaas:).tw. (3599)
3. lovaas-o:.au. (2)
4. or/1-3 (4011)
5. exp behavior therapy/ (9758)
6. "Early Intervention (Education)"/ (450)
7. (early adj1 intervention:).tw. (2550)
8. (intense: or intensive or discreet: or discrete:).tw. (70017)
9. trial learning.tw. (43)
10. (applied adj1 behavio: adj1 analysi:).tw. (50)
11. (young autis: adj1 (project: or study or studies)).tw. (0)
12. or/5-11 (82168)
13. 4 and 12 (294)
14. limit 13 to yr=1998-2004 (253)
15. prospective studies/ (102975)
16. (prospective adj1 (study or studies or trial:)).tw. (33273)
17. multivariate analysis/ (23950)
18. (multivariate adj1 analy:).tw. (24322)
19. risk factors/ (154898)
20. (risk adj1 factor:).tw. (81275)
21. odds ratio/ (17062)
22. (odds adj1 ratio:).tw. (30740)
23. evaluation studies/ (20379)
24. (evaluation adj1 (study or studies or trial:)).tw. (2392)
25. (comparison adj1 group adj1 design).tw. (64)
26. (historical adj1 cohort:).tw. (574)
27. clinical trials/ (33873)
28. clinical trials, phase I/ (1507)
29. clinical trials, phase II/ (1948)
30. clinical trials, phase III/ (1689)
31. clinical trials, phase IV/ (59)
32. clinical trial.pt. (189180)
33. clinical trial, phase I.pt. (4649)
34. clinical trial, phase II.pt. (7200)
35. clinical trial, phase III.pt. (2130)
36. clinical trial, phase IV.pt. (158)
37. controlled clinical trial.pt. (21492)
38. ((clinical or controlled) adj1 (study or studies or trial:)).tw. (100740)
39. multicenter study.pt. (42987)
40. multicenter studies/ (4599)
41. ((multicent: or multi-cent:) adj1 (study or studies or trial:)).tw. (97141)
42. comparative study/ (418815)
43. (comparative adj1 (study or studies or trial:)).tw. (13252)
44 sampling studies/ (6968)
45 (sampling adj1 (study or studies or trial:)).tw. (212)
46 program evaluation/ (14344)
47 ([program: or programme:)] adj2 evaluat:].tw. (3475)
48 case-control studies/ (49035)
49 (case adj1 control).tw. (17950)
50 research/ (23714)
51 research design/ (17704)
52 (research adj2 design:).tw. (5983)
53 or/15-52 (966385)
54 14 and 53 (78)
55 cohort studies/ (38869)
56 cohort effect/ (215)
57 longitudinal studies/ (20569)
58 follow-up studies/ (123746)
59 prospective studies/ (102975)
60 retrospective studies/ (121461)
61 cross-sectional studies/ (35789)
62 (cohort adj1 (study or studies or analys: or trial: or effect:)).tw. (17861)
63 (longitudinal adj1 (study or studies or analys: or trial:)).tw. (10983)
64 ((follow-up or follow) adj1 (study or studies or analys: or trial:)).tw. (10551)
65 (prospective adj1 (study or studies or analys: or trial:)).tw. (34509)
66 (retrospective adj1 (study or studies or analys: or trial:)).tw. (29915)
67 (concurrent adj1 (study or studies or analys: or trial:)).tw. (384)
68 (incidence adj1 (study or studies or analys: or trial:)).tw. (2569)
69 (cross-sectional adj1 (study or studies or analys: or trial:)).tw. (13206)
70 (case adj1 series).tw. (6369)
71 (case adj1 (study or studies or analys: or trial:)).tw. (13083)
72 (case adj1 design).tw. (2377)
73 (time adj1 series).tw. (2972)
74 multivariate analysis/ (23950)
75 (multivariate adj1 analys:).tw. (24322)
76 risk factors/ (154898)
77 (risk adj1 factor:).tw. (81275)
78 odds ratio/ (17062)
79 (odds adj1 ratio:).tw. (30740)
80 survival analysis/ (34372)
81 (survival adj2 (study or studies or analys:)).tw. (9910)
82 ((uncontrol: or un-control:) adj1 (study or studies or analys: or trial:)).tw. (719)
83 ((non-control: or noncontrol:) adj1 (study or studies or analys: or trial:)).tw. (94)
84 (observational adj1 (study or studies or trial:)).tw. (7605)
85 research/ (23714)
86 research design/ (17704)
87 (research adj2 design:).tw. (5983)
88 or/55-87 (628978)
89 14 and 88 (55)
90 54 or 89 (101)
91 limit 90 to human (100)
EMBASE (OVID)
1998 to 2004 Week 47
Database: EMBASE <1996 to 2004 Week 50>
Search Strategy:

1 exp Autism/ (4484)
2 (autis: or asperger: or kanner: or lovaas:).tw. (3854)
3 lovaas-o:.au. (3)
4 or/1-3 (4802)
5 behavior therapy/ (9060)
6 (early adj1 intervention).tw. (2348)
7 (intense: or intensive or discreet: or discrete:).tw. (70033)
8 trial learning.tw. (38)
9 (applied adj1 behavior: adj1 analytical).tw. (47)
10 (young autis: adj1 (project: or study or studies)).tw. (2)
11 or/5-10 (81033)
12 4 and 11 (292)
13 limit 12 to yr=1998-2004 (260)
14 prospective study/ (34972)
15 (prospective adj1 (study or studies or trial:)).tw. (34460)
16 multivariate analysis/ (16941)
17 (multivariate adj1 analysis:).tw. (23787)
18 risk factor/ (105967)
19 (risk adj1 factor:).tw. (79681)
20 exp risk/ (251850)
21 (odds adj1 ratio:).tw. (29059)
22 (comparison adj1 group adj1 design:).tw. (74)
23 (historical adj1 cohort:).tw. (551)
24 follow up/ (113807)
25 (evaluation adj1 (study or studies or trial:)).tw. (2524)
26 clinical trial/ (249654)
27 phase 1 clinical trial/ (6773)
28 phase 2 clinical trial/ (9972)
29 phase 3 clinical trial/ (4764)
30 phase 4 clinical trial/ (281)
31 controlled study/ (1304272)
32 ((clinical or controlled) adj1 (study or studies or trial:)).tw. (103435)
33 multicenter study/ (25867)
34 ((multicent: or multi-cent:) adj1 (study or studies or trial:)).tw. (10638)
35 comparative study/ (40047)
36 (comparative adj1 (study or studies or trial:)).tw. (13418)
37 (sampling adj1 (study or studies or trial:)).tw. (233)
38 ((program: or programme: adj2 evaluation):.tw. (3192)
39 case-control study/ (8492)
40 (case adj1 control).tw. (17605)
41 methodology/ (32043)
42 (research adj2 design:).tw. (4768)
43 or/14-42 (1744731)
13 and 43 (111)
44 cohort analysis/ (20830)
45 epidemiology/ (3428)
46 longitudinal study/ (7919)
47 follow up/ (113807)
48 prospective study/ (34972)
49 retrospective study/ (43761)
50 prevalence/ (62309)
51 (cohort adj (study or studies or analys: or trial: or effect:)).tw. (17995)
52 (longitudinal adj (study or studies or analys: or trial:)).tw. (10202)
53 (follow-up or followup) adj (study or studies or analys: or trial:).tw. (10716)
54 (prospective adj (study or studies or analys: or trial:)).tw. (35788)
55 (retrospective adj (study or studies or analys: or trial:)).tw. (29421)
56 (concurrent adj (study or studies or analys: or trial:)).tw. (386)
57 (incidence adj (study or studies or analys: or trial:)).tw. (2429)
58 ((cross-sectional or crosssectional) adj (study or studies or analys: or trial:)).tw. (23787)
59 risk/ (11296)
60 (risk adj factor:).tw. (79681)
61 (odds adj ratio:).tw. (29059)
62 exp survival/ (101584)
63 (survival adj2 (study or studies or analys:)).tw. (9675)
64 ((uncontrol: or un-control:) adj (study or studies or analys: or trial:)).tw. (801)
65 (non-control: or noncontrol:) adj (study or studies or analys: or trial:).tw. (111)
66 (observation adj (study or studies or trial:)).tw. (7805)
67 research/ (11584)
68 methodology/ (32043)
69 (research adj2 design:).tw. (4768)
70 or/45-77 (532496)
71 13 and 78 (63)
72 44 or 79 (130)
73 limit 80 to human (125)
74 limit 81 to (infant <to one year> or child <unspecified age> or preschool child <1 to 6 years> or school child <7 to 12 years> or adolescent <13 to 17 years>) (80)

213.
48 follow-up studies/ (34134)
49 prospective studies/ (34134)
50 retrospective studies/ (0)
51 cross-sectional studies/ (11503)
52 (cohort adj1 (study or studies or analys: or trial: or
effect:)).tw. (3371)
53 (longitudinal adj1 (study or studies or analys: or trial:)).tw.
(2605)
54 ((follow-up or followup) adj1 (study or studies or analys: or
trial:)).tw. (1600)
55 (prospective adj1 (study or studies or analys: or trial:)).tw. (3950)
56 (retrospective adj1 (study or studies or analys:)).tw. (2571)
57 (concurrent adj1 (study or studies or analys: or trial:)).tw. (63)
58 (incidence adj1 (study or studies or analys: or trial:)).tw. (284)
59 (cross-sectional adj1 (study or studies or analys: or
trial:)).tw. (2332)
60 (case adj1 series).tw. (662)
61 (case adj1 (study or studies or analys: or trial:)).tw. (9155)
62 (case adj1 design).tw. (345)
63 (time adj1 series).tw. (359)
64 multivariate analysis/ (4170)
65 (multivariate adj1 analys:).tw. (2172)
66 risk factors/ (17553)
67 (risk adj1 factor:).tw. (12040)
68 odds ratio/ (14365)
69 (odds adj1 ratio:).tw. (4184)
70 survival analysis/ (2252)
71 (survival adj2 (study or studies or analys:)).tw. (513)
72 ((uncontrol: or un-control:) adj1 (study or studies or analys:
or
trial:)).tw. (94)
73 ((non-control: or noncontrol:) adj1 (study or studies or analys:
or trial:)).tw. (14)
74 (observational adj1 (study or studies or trial:)).tw. (1346)
75 research/ (7520)
76 research design/ (1416)
77 (research adj2 design:).tw. (3655)
78 or/1-77 (100540)
79 15 and 78 (23)
80 44 or 79 (34)

Psyclnfo (OVID)
Database: PsyclNFO <1985 to November Week 5 2004>
Search Strategy:

1 exp autism/ (6032)
2 aspergers syndrome/ (617)
3 autistic children/ (2279)
4 autistic thinking/ (26)
5 (autis: or asperger: or kanner: or lovaas:).tw. (9053)
6 lovaas-o:.au. (25)
7 or/1-6 (9115)
8 exp behavior therapy/ (6635)
9 early intervention/ (3899)
10 [early adj1 intervention:].tw. (3605)
11 (intense: or intensive or discreet: or discrete:).tw. (17396)
trial learning.tw. (98)
(applied adj1 behavio: adj1 analys:).tw. (536)
(young autis: adj1 (project: or study or studies)).tw. (12)
or/8-14 (2953)
7 and 15 (640)
prospective studies/ (268)
(prospective adj1 (study or studies or trial:)).tw. (3757)
exp multivariate analysis/ (5219)
(multivariate adj1 analys:).tw. (5014)
risk factors/ (5797)
(risk adj1 factor:).tw. (16399)
(odds adj1 ratio:).tw. (1371)
(evaluation adj1 (study or studies or trial:)).tw. (1035)
(comparison adj1 group adj1 design).tw. (91)
(historical adj1 cohort:).tw. (41)
((clinical or controlled) adj1 (study or studies or trial:)).tw. (14134)
((multicent: or multi-cent:) adj1 (study or studies or trial:)).tw. (761)
(comparative adj1 (study or studies or trial:)).tw. (3882)
(sampling adj1 (study or studies or trial:)).tw. (126)
(case adj1 control).tw. (1441)
exp experimentation/ (21245)
(research adj2 design:).tw. (5855)
or/17-33 (76695)
16 and 34 (40)
cohort analysis/ (581)
longitudinal studies/ (11806)
retropective studies/ (264)
(cohort adj1 (study or studies or analys: or trial: or effect:)).tw. (1896)
(longitudinal adj1 (study or studies or analys: or trial:)).tw. (14295)
((follow-up or followup) adj1 (study or studies or analys: or trial:)).tw. (5604)
(prospective adj1 (study or studies or analys: or trial:)).tw. (4025)
(retrospective adj1 (study or studies or analys:)).tw. (2172)
(concurrent adj1 (study or studies or analys: or trial:)).tw. (176)
(incidence adj1 (study or studies or analys: or trial:)).tw. (284)
(cross-sectional adj1 (study or studies or analys: or trial:)).tw. (2450)
(case adj1 series).tw. (596)
(case adj1 (study or studies or analys: or trial:)).tw. (22730)
(case adj1 design:).tw. (252)
(time adj1 series).tw. (1696)
(multivariate adj1 analys:).tw. (5014)
risk factors/ (5797)
(risk adj1 factor:).tw. (16399)
(odds adj1 ratio:).tw. (1371)
(survival adj2 (study or studies or analys:)).tw. (794)
((uncontrol: or un-control:) adj1 (study or studies or analys: or trial:)).tw. (134)
((non-control: or noncontrol:) adj1 (study or studies or analys: or trial:)).tw. (13)
(observational adj1 (study or studies or trial:)).tw. (963)
exp experimentation/ (21245)
(research adj2 design::).tw. (5855)
or/36-60 (106632)
16 and 61 (85)
limit 62 to yr=1998 - 2004 (50)
limit 63 to ((100 childhood <birth to age 12 yrs> or 200 adolescence <age 13 to 17 yrs>) and general public) (0)
(child: or teen: or adolescent: or youth or infant: or baby or babies).tw. (259612)
63 and 65 (44)

Additional Electronic Databases
- Academic Search Premier
- Dissertation Abstracts
- HTA database: Centre for Reviews and Dissemination, University of York, UK.
- PapersFirst

Conference Proceedings
- ProceedingsFirst

Current Awareness Service
- Web of Science

Fugitive Literature Search
- HSTAT (National Library of Medicine link to full-text documents: Health Services/Technology Assessment text)
- LocatorPlus
- National Library of Medicine (NLM Gateway)
- WorldCat

Other Sources
- Reference lists of retrieved studies
- Internet web sites
### Personal Communication

<table>
<thead>
<tr>
<th>Contact</th>
<th>Institution</th>
<th>Date</th>
<th>Citations retrieved</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scott Cross</td>
<td>Lovaas Institute</td>
<td>March 2005</td>
<td>Eikeseth, et al. 2002 Cohen, et al., 2006 Sallows &amp; Graupner, 2005</td>
<td>N/A</td>
</tr>
<tr>
<td>Carolyn Green</td>
<td>University of Victoria</td>
<td>January 2006</td>
<td>Eikeseth, et al., in press</td>
<td>N/A</td>
</tr>
<tr>
<td>Tristram Smith</td>
<td>University of Rochester</td>
<td>January 2007</td>
<td>No new citations</td>
<td>N/A</td>
</tr>
<tr>
<td>Glen Sallows</td>
<td>Wisconsin Early Autism Project, Madison, WI</td>
<td>January 2007</td>
<td>N/A</td>
<td>No reply</td>
</tr>
<tr>
<td>Tamlynn Graupner</td>
<td>Wisconsin Early Autism Project, Madison, WI</td>
<td>January 2007</td>
<td>N/A</td>
<td>No reply</td>
</tr>
<tr>
<td>Mila Amerine-</td>
<td>Central Valley Autism Project, Modesto, CA</td>
<td>January 2007</td>
<td>N/A</td>
<td>Referred to Tristram Smith</td>
</tr>
<tr>
<td>Dickens</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jane Howard</td>
<td>California State University, Stanislaus and The Kendall school</td>
<td>January 2007</td>
<td>N/A</td>
<td>No reply</td>
</tr>
<tr>
<td>Svein Eikeseth</td>
<td>Akershus University College, Norway</td>
<td>January 2007</td>
<td>N/A</td>
<td>No reply</td>
</tr>
</tbody>
</table>
### Appendix F

#### Study Selection Form

**Systematic Review of Lovaas Therapy**

<table>
<thead>
<tr>
<th>Reviewer:</th>
<th>Date reviewed:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study Title:</th>
<th>Year:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author(s):</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Citation:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

#### Level of Review (please check): Title: Abstract: Full Article: 

<table>
<thead>
<tr>
<th>Action:</th>
<th>Include</th>
<th>Obtain for reference/background info (Ref-back)</th>
<th>Exclude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Exclusion Code(s) (see key):** 

**Review Question:** How much improvement in intellectual, linguistic and social functioning do autistic children under the age of 6 achieve with Lovaas-type applied behavioural analysis therapy?

<table>
<thead>
<tr>
<th>✓</th>
<th>Selection Criteria</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population</td>
<td>Autistic children &lt; 6 years</td>
<td>All others</td>
</tr>
<tr>
<td></td>
<td>Interventions</td>
<td>Lovaas-type ABA compared to another intervention</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outcome measures</td>
<td>Intellectual functioning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study designs</td>
<td>Utilizing treatment and control groups:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Languages</td>
<td>All</td>
<td></td>
</tr>
</tbody>
</table>

- Early: Initiated before age 6
- Intensive: 20 - 40 hours/week
- Intellectual functioning
- Language
- Social interaction and play
- Adaptive or self-care skills
- Maladaptive behaviour
- Randomized Controlled Trials
- Cohort studies
- Case control studies
- Systematic reviews
- Case reports
- Case series
- None
<table>
<thead>
<tr>
<th>Codes*</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Include</td>
<td>Obtain full paper to assess</td>
</tr>
<tr>
<td>Ref-back</td>
<td>Obtain for reference or background information</td>
</tr>
<tr>
<td>Exclusions</td>
<td></td>
</tr>
<tr>
<td>ECL</td>
<td>Editorial, comment, or letter</td>
</tr>
<tr>
<td>NR-design/methods</td>
<td>Not relevant: design or methodological issues</td>
</tr>
<tr>
<td>NR-IA</td>
<td>Not relevant: implementation/application issues</td>
</tr>
<tr>
<td>NR-OCD</td>
<td>Not relevant: opinion/commentary/description</td>
</tr>
<tr>
<td>NR-ROS</td>
<td>Not relevant: report of study</td>
</tr>
<tr>
<td>NR-Review</td>
<td>Not relevant: review/overview</td>
</tr>
<tr>
<td>NR-Stat Meth</td>
<td>Not relevant: statistical methodology issues</td>
</tr>
<tr>
<td>NR- Other</td>
<td>Not relevant: other reasons/issues</td>
</tr>
<tr>
<td>NR- Text word only</td>
<td>Not relevant: studies identified had text word in title and/or abstract, but not relevant content.</td>
</tr>
</tbody>
</table>

Appendix G

Study Quality Assessment Tool for RCTs: Modified Jadad Scale

Instructions

Please read each research article and answer the following questions. It should not take more than 10 minutes to score each article. There are no right or wrong answers.

(Note: Font "strikethrough" is displayed to show how the Modified Jadad Scale was revised to fit a quality assessment of behavioural research).

Scoring the items

Assign a score of 1 point for each yes response or 0 points for each no response to the following questions. (There are no in-between marks)

1. Was the study described as randomized? (This includes the use of words such as randomly, random, and randomization).
2. Was the study described as double blind? For this thesis, the question was revised to: Was the outcome assessor blinded?
3. Was there a clear description of the inclusion/exclusion criteria?
4. Was the method used to assess adverse effects described? (Not applicable for this thesis)
5. Were the methods of statistical analysis described?
6. Was there a description of withdrawals and dropouts?

Give 1 additional point if:

For question 1, the method to generate the sequence of randomization was described AND it was APPROPRIATE (e.g., table of random numbers, computer generated, etc.)

And/or, if for question 2, the method of double blinding was described AND it was APPROPRIATE (e.g., identical placebo, active placebo, dummy, etc).

Deduct 1 point if:

For question 1, the method to generate the sequence of randomization was described AND it was INAPPROPRIATE (e.g., patients were allocated alternately, or according to date of birth, hospital number, etc). 

And/or, if for question 2, the study was described as double blind but the method of blinding was INAPPROPRIATE (e.g., comparison of tablet vs. injection with no double dummy).

Guidelines for Assessment

Randomization

A method to generate the sequence of randomization will be regarded as appropriate if it allowed each study participant to have the same chance of receiving each intervention and the investigators could not predict which treatment was next. Methods of allocation using date of birth, date of admission hospital numbers, or alternation should be not regarded as appropriate.

Double blinding

A study must be regarded as double blind if the word "double blind" is used. The method will be regarded as appropriate if it is stated that neither the person doing the assessments nor the study participant could not identify the intervention being assessed, or if in the absence of such a statement the use of active placebos, identical placebos, or dummies is mentioned.

Withdrawals and Dropouts

Participants who were included in the study but did not complete the observation period or who were not included in the analysis must be described. The number AND the reasons for withdrawal in each group must be stated. If there were no withdrawals, it should be stated in the article. If there is no statement on withdrawals, this item must be given no points.
# Study Quality Assessment Tool for RCTs

## Modified Jadad Scale Worksheet

**Reviewer:** ______________________  **Date reviewed:** ______________________

**Study Name:** ______________________

**Authors:** ______________________  **Year:** ______________________

**Journal:** ______________________

<table>
<thead>
<tr>
<th>Item</th>
<th>Basic Score</th>
<th>PLUS: Additional Points</th>
<th>MINUS: Deducted Points</th>
<th>Final Item Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the study described as <strong>randomized</strong>? (This includes the use of words such as randomly, random, and randomization).</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Was the outcome <strong>assessor blinded</strong>?</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Was there a clear description of the <strong>inclusion/exclusion</strong> criteria?</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Were the methods of <strong>statistical analysis</strong> described?</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Was there a description of <strong>withdrawals and dropouts</strong>?</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**OVERALL QUALITY SCORE:** /7

**Reviewer Notes:**

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________

1 Jadad, et al (1996) – Further revisions made by this author to match design elements of behavioural research (indicated by "strike-through" text).
Appendix H
Study Quality Assessment Tool for Observational Studies: Reisch et al.

Reviewer: ______________________  Date: ______________________

JOURNAL: ______________________
Volume: _____  Number: _____  Pages:  _______  Year: _______

AUTHOR(S): ______________________

TITLE: ______________________

"Y"=Yes; "N"=No; "U"=Unclear or Unknown; "NA"=Not applicable; "T/M"=Treatment or Management Method.
Note: * indicates those items considered ESSENTIAL to a well-executed clinical trial.

<table>
<thead>
<tr>
<th>1. PURPOSE OF STUDY</th>
<th>3. SAMPLE SIZE DETERMINATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Title consistent with purpose of the study?</td>
<td>Y  N  U</td>
</tr>
<tr>
<td>1. Method</td>
<td>Y*  N  U  NA</td>
</tr>
<tr>
<td>a. number of subjects predetermined; OR: b. sequential experimental design used.</td>
<td></td>
</tr>
<tr>
<td>B. Statement of purpose given?</td>
<td>Y*  N  U</td>
</tr>
<tr>
<td>2. Statement made that time period of study predetermined</td>
<td>Y  N  U</td>
</tr>
<tr>
<td>C. Endpoints of therapeutic effects specified?</td>
<td>Y*  N  U</td>
</tr>
<tr>
<td>3. Specified time period from to (fill in period)</td>
<td>Y  N  U</td>
</tr>
<tr>
<td>D. Magnitude of T/M effect under investigation specified?</td>
<td>Y*  N  U</td>
</tr>
<tr>
<td>4. No method specified (check if applicable)</td>
<td>-  -  -</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. EXPERIMENTAL DESIGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Data Collection (Circle only one number)</td>
</tr>
<tr>
<td>1* Planned prior to T/M of subjects; data collected prospectively under specified conditions.</td>
</tr>
<tr>
<td>2. Planned prior to T/M of subjects; data collected retrospectively by record review.</td>
</tr>
<tr>
<td>3. Not planned prior to T/M of subjects; data collected retrospectively.</td>
</tr>
<tr>
<td>4. Unclear time relation to T/M to data collection.</td>
</tr>
<tr>
<td>B. Total # of subjects specified</td>
</tr>
<tr>
<td>Total number of subjects is:</td>
</tr>
<tr>
<td>(Do not answer 3C and 3D if the magnitude of the T/M difference is not specified)</td>
</tr>
<tr>
<td>C. Adequate # of subjects planned to detect magnitude of T/M differences under investigation.</td>
</tr>
<tr>
<td>Y  N  U</td>
</tr>
<tr>
<td>D. Adequate # of subjects enrolled to detect magnitude of T/M differences under investigation.</td>
</tr>
<tr>
<td>Y*  N  U</td>
</tr>
</tbody>
</table>
### 4. DESCRIPTION AND SUITABILITY OF SUBJECTS

A. Description of subjects.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of subjects given.</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>Race of subjects given.</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>Sex of subjects given.</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
</tbody>
</table>

B. Studies with data collected prospectively.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y*</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

C. Subjects selected for this study suitable for question(s) posed by these researchers.

<table>
<thead>
<tr>
<th>Subjects selected for this study suitable for question(s) posed by these researchers.</th>
<th>Y*</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

### 5. RANDOMIZATION TO T/M GROUPS

A. It is possible to design a randomized study to evaluate the T/M under consideration.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

B. Randomization claimed and documented.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y*</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

C. Randomization not performed and bias is likely.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

### 6. STRATIFICATION (cont)

A. Analysis of data based on stratification of groups into homogeneous subgroups.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y*</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

B. Studies with data collected retrospectively.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

C. Subjects matched/paired but assignment to T/M groups not randomized.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

D. Subject as own control but T/M order not randomized.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

E. Subjects compared according to their response to the T/M procedure.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

F. Convenience (subjects selected for availability).

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

G. Control (comparison) group not included.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

H. Other non-random (explain):

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

### 6. STRATIFICATION (cont)

A. Description of subjects.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

B. Studies with data collected retrospectively.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y*</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

C. Subjects matched/paired but assignment to T/M groups not randomized.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

D. Subject as own control but T/M order not randomized.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

E. Subjects compared according to their response to the T/M procedure.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

F. Convenience (subjects selected for availability).

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

G. Control (comparison) group not included.

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
</table>

H. Other non-random (explain):

<table>
<thead>
<tr>
<th>Subjects given.</th>
<th>Y</th>
<th>N</th>
<th>U</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. PROCEDURES FOR T/M</td>
<td>9. BLINDING (Cont)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Clear specification of:</td>
<td>B. Blinding claimed but nature of T/M makes claim unrealistic.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Dosage</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>2. Time of day administered</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>3. Frequency</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>4. Total duration of T/M</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>5. Route (IV, IM, PO, etc)</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>6. Presentation (tablet, syrup, etc)</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>7. Source for drug or equipment in T/M under investigation</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>8. Indications for:</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>a. Initiation of T/M</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>b. Modification of T/M</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
<td>NA*</td>
</tr>
<tr>
<td>c. Discontinuation of T/M</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
<td>NA*</td>
</tr>
<tr>
<td>B. Carry-over or refractory effects avoided or allowed for in the design of the study</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>C. Description of all subjects or their records which were lost or dropped</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>D. Loss of subjects or their records likely to bias the results of this study</td>
<td>Y</td>
<td>N*</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>10. SUBJECT ATTENTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Specific procedures established to minimize loss of subjects from this study</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>B. Subjects or their entire records lost/dropped</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>C. Description of all subjects or their records which were lost or dropped</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>D. Loss of subjects or their records likely to bias the results of this study</td>
<td>Y</td>
<td>N*</td>
<td>U</td>
<td>NA</td>
</tr>
<tr>
<td>11. PRESENTATION OF DATA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Text clearly understandable</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>B. Explanation given for any variation in numbers of subjects in same T/M groups for different comparisons</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
<td>NA*</td>
</tr>
<tr>
<td>C. Clear identification of reported variables</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>D. Information given to support text, figure, table</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>E. Computation errors change the interpretation of results for any important variable</td>
<td>Y</td>
<td>N*</td>
<td>U</td>
<td>NA*</td>
</tr>
<tr>
<td>F. Table, figure or text contradicted</td>
<td>Y</td>
<td>N</td>
<td>U</td>
<td>NA</td>
</tr>
</tbody>
</table>

* A variable is important only when it is clearly identified by the author(s) in the abstract or in the statement of purpose to describe differences between groups related to their treatment or management.
<table>
<thead>
<tr>
<th>11. PRESENTATION OF DATA (Cont)</th>
<th>Y</th>
<th>N</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>G. Tables and figures referenced in text.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>12. ANALYSIS OF DATA</th>
<th>Y*</th>
<th>N</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Descriptive measures (mean, range, standard deviation, proportion, etc) given for all important variables.</td>
<td>All</td>
<td>Some</td>
<td>None</td>
</tr>
<tr>
<td>B. Statistical tests used for comparisons involving important variables.</td>
<td>All</td>
<td>Some</td>
<td>None</td>
</tr>
<tr>
<td>C. Reported statistical tests are:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Clearly identified</td>
<td>All</td>
<td>Some</td>
<td>None</td>
</tr>
<tr>
<td>2. Appropriately used</td>
<td>All</td>
<td>Some</td>
<td>None</td>
</tr>
<tr>
<td>3. Appropriately interpreted.</td>
<td>All</td>
<td>Some</td>
<td>None</td>
</tr>
<tr>
<td>D. Responses to items 12B, C1, C2, C3 marked &quot;ALL.&quot;</td>
<td>Y*</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>13. EVALUATION OF SUBJECTS AND T/M</th>
<th>Y*</th>
<th>N</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. All important clinical information reported. If no or unclear, explain:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Laboratory and other measurements appear standardized and consistent.</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
</tr>
<tr>
<td>C. Subject compliance assessed.</td>
<td>Y*</td>
<td>N</td>
<td>U NA</td>
</tr>
<tr>
<td>D. Evaluation methods adequately described.</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
</tr>
<tr>
<td>E. Evaluation methods appropriate to answer question(s) posed by investigators.</td>
<td>Y*</td>
<td>N</td>
<td>U NA</td>
</tr>
<tr>
<td>F. Evaluation of potential hazards (Circle only one)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*. Prospective evaluation of all likely hazard(s).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Prospective evaluation of some hazard(s).</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14. RECOMMENDATIONS</th>
<th>Y*</th>
<th>N</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Recommendation(s) are: (Circle one)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Recommendation for use of T/M method based on a controlled randomized prospective study; made only if convincing benefit is demonstrated and all important hazards assessed; and applied to subjects and conditions similar to those in this study.</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
</tr>
<tr>
<td>2. Recommendation against use of T/M method supported by data relating to hazards or toxicity of T/M or supported by calculation of appropriate confidence intervals.</td>
<td>Y*</td>
<td>N</td>
<td>U NA</td>
</tr>
<tr>
<td>3. Recommendations neither for nor against use of a T/M method is appropriate since criteria in 14B1 and 14B2 are not met.</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14. CONCLUSIONS</th>
<th>Y*</th>
<th>N</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Recommendation(s) are: (Circle one)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Recommendation for use of T/M method based on a controlled randomized prospective study; made only if convincing benefit is demonstrated and all important hazards assessed; and applied to subjects and conditions similar to those in this study.</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
</tr>
<tr>
<td>2. Recommendation against use of T/M method supported by data relating to hazards or toxicity of T/M or supported by calculation of appropriate confidence intervals.</td>
<td>Y*</td>
<td>N</td>
<td>U NA</td>
</tr>
<tr>
<td>3. Recommendations neither for nor against use of a T/M method is appropriate since criteria in 14B1 and 14B2 are not met.</td>
<td>Y*</td>
<td>N</td>
<td>U</td>
</tr>
</tbody>
</table>
18. RATINGS FOR PAPER

Rating 1 is a subjective rating with each section rated on the following basis: "2" indicates superior performance on a section; "1" indicates undistinguished performance on a section; and "0" indicates poor performance. Rating 1 is the ratio of the total to 28.

Rating 2 is a more objective rating assessment on the basis of starred items*. Rating 2 is the ratio of the starred items chosen by the reviewer to the maximum possible total.

<table>
<thead>
<tr>
<th>QUALITY INDICATOR</th>
<th>SYNOPSIS OF ITEMS REVIEWED</th>
<th># of starred items (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item #</td>
<td>Rating 1</td>
<td>Rating 2</td>
</tr>
<tr>
<td>PURPOSE OF STUDY</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>EXPERIMENTAL DESIGN</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>SAMPLE SIZE DETERMINATION</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DESCRIPTION AND SUITABILITY OF SUBJECTS</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>RANDOMIZATION TO T/M</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>STRATIFICATION</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>CONTROL USAGE</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>PROCEDURES FOR TREATMENT/MANAGEMENT</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>BLINDING (MASKING)</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>SUBJECT ATTRITION</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>PRESENTATION OF DATA</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>ANALYSIS OF DATA</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>EVALUATION OF SUBJECTS AND T/M</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>RECOMMENDATIONS / CONCLUSIONS</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum Possible Total</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Ratio of total to total maximum possible</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The maximum possible total for Rating 2 is determined by subtracting the Total "NA" responses marked by the reviewer from 33. As many as 7 "NA" responses may be recorded (Section 8, 11, 12, 14, 16).

++ Maximum Possible Total (Rating 2) = 33 minus Number of "NA" items selected. (NA* = )

*Starred items are those considered essential to a well-executed clinical trial.

CITATION:
Appendix I
Data Extraction Form

Systematic Review of Lovaas Therapy
Experimental/Observational Studies

Reviewer: ____________________________ Date reviewed: ____________________________

Study Title: ____________________________

Author(s): ____________________________ Year: ____________________________

Source: ____________________________

Institutional Affiliation (first author) and/or address: ____________________________

Notes: ____________________________

Review Question: How much improvement in intellectual, linguistic and social functioning do autistic children under the age of 6 achieve with Lovaas-type applied behavioural analysis therapy?

<table>
<thead>
<tr>
<th>Items &amp; Guidelines</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design hierarchy</td>
<td>EXPERIMENTAL/QUASI-EXPERIMENTAL:</td>
</tr>
<tr>
<td></td>
<td>• Randomized Controlled Trial</td>
</tr>
<tr>
<td></td>
<td>• Controlled Trial</td>
</tr>
<tr>
<td></td>
<td>OBSERVATIONAL:</td>
</tr>
<tr>
<td></td>
<td>• Cohort study with matched concurrent controls</td>
</tr>
<tr>
<td></td>
<td>• Cohort study with unmatched concurrent controls</td>
</tr>
<tr>
<td></td>
<td>• Cohort study with historic controls</td>
</tr>
<tr>
<td></td>
<td>• Case Control</td>
</tr>
<tr>
<td>General Comments:</td>
<td></td>
</tr>
<tr>
<td>Study question</td>
<td></td>
</tr>
<tr>
<td>Inclusion criteria</td>
<td></td>
</tr>
<tr>
<td>• Specified and replicable?</td>
<td></td>
</tr>
<tr>
<td>Match the goals of the study?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Autistic children &lt; 6 years [if not, exclude study]</td>
</tr>
<tr>
<td></td>
<td>• Autism diagnosed as per DSM criteria?</td>
</tr>
<tr>
<td>General comments:</td>
<td></td>
</tr>
<tr>
<td>Exclusion criteria</td>
<td></td>
</tr>
<tr>
<td>• Specified and replicable?</td>
<td></td>
</tr>
<tr>
<td>• Match the goals of the study?</td>
<td></td>
</tr>
<tr>
<td>General comments:</td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td></td>
</tr>
<tr>
<td>• Clearly described?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Sex: Males: n= Females: n=</td>
</tr>
<tr>
<td>Items &amp; Guidelines</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Participants</td>
<td></td>
</tr>
<tr>
<td>(continued)</td>
<td></td>
</tr>
<tr>
<td>• Sufficient #?</td>
<td></td>
</tr>
<tr>
<td>• All accounted for?</td>
<td></td>
</tr>
<tr>
<td>• Baseline</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>comparability of the groups documented?</td>
<td></td>
</tr>
<tr>
<td>• DSM-IV diagnostic criteria?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Age (mean and range) Treatment group: mean: range:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ Control group(s): mean: range:</td>
</tr>
<tr>
<td>□ Ethnicity (specify):</td>
<td></td>
</tr>
<tr>
<td>□ Treatment group: n=</td>
<td></td>
</tr>
<tr>
<td>□ Treatment group(s): n=</td>
<td></td>
</tr>
<tr>
<td>□ Geographical regions:</td>
<td></td>
</tr>
<tr>
<td>□ Recruitment procedures:</td>
<td></td>
</tr>
<tr>
<td>□ Diagnostic criteria:</td>
<td></td>
</tr>
<tr>
<td>General comments:</td>
<td></td>
</tr>
<tr>
<td>Items &amp; Guidelines</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Interventions</strong></td>
<td>[ ] Lovaas autism treatment vs control (if not, exclude study)</td>
</tr>
<tr>
<td></td>
<td>[ ] Name of control intervention(s):</td>
</tr>
<tr>
<td></td>
<td>Control #1:</td>
</tr>
<tr>
<td></td>
<td>Control #2:</td>
</tr>
<tr>
<td></td>
<td>[ ] Treatment: N=</td>
</tr>
<tr>
<td></td>
<td>Control #1; N=</td>
</tr>
<tr>
<td></td>
<td>Control #2; N=</td>
</tr>
<tr>
<td></td>
<td>[ ] Setting of each intervention (ie: clinic or home):</td>
</tr>
<tr>
<td></td>
<td>Treatment:</td>
</tr>
<tr>
<td></td>
<td>Control #1:</td>
</tr>
<tr>
<td></td>
<td>Control #2:</td>
</tr>
<tr>
<td></td>
<td>[ ] Person administering each intervention:</td>
</tr>
<tr>
<td></td>
<td>Treatment:</td>
</tr>
<tr>
<td></td>
<td>Control #1:</td>
</tr>
<tr>
<td></td>
<td>Control #2:</td>
</tr>
<tr>
<td></td>
<td>[ ] Ratio of therapist to child(ren):</td>
</tr>
<tr>
<td></td>
<td>Treatment:</td>
</tr>
<tr>
<td></td>
<td>Control #1:</td>
</tr>
<tr>
<td></td>
<td>Control #2:</td>
</tr>
<tr>
<td></td>
<td>[ ] Staff training provided?:</td>
</tr>
<tr>
<td></td>
<td>Treatment:</td>
</tr>
<tr>
<td></td>
<td>Control #1:</td>
</tr>
<tr>
<td></td>
<td>Control #2:</td>
</tr>
<tr>
<td>Items &amp; Guidelines (continued)</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------</td>
</tr>
</tbody>
</table>
| □ Intensity of each program (average # hrs/week): | Treatment:  
Control #1:  
Control #2: |
| □ Duration of each program: | Treatment:  
Control #1:  
Control #2: |
| □ Length of post-intervention follow-up (week/months): | Treatment:  
Control #1:  
Control #2: |
| □ Were treatment and control groups comparable?: |

General comments:
<table>
<thead>
<tr>
<th>Methods</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Randomization</td>
<td></td>
</tr>
<tr>
<td>• Allocation concealment</td>
<td></td>
</tr>
<tr>
<td>• Study duration</td>
<td></td>
</tr>
<tr>
<td>• Blinding</td>
<td></td>
</tr>
<tr>
<td>• Drop outs/crossovers</td>
<td></td>
</tr>
<tr>
<td>• Confounders</td>
<td></td>
</tr>
<tr>
<td>• Co-interventions:</td>
<td></td>
</tr>
</tbody>
</table>

General comments:
<table>
<thead>
<tr>
<th>Items &amp; Guidelines</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes</td>
<td></td>
</tr>
<tr>
<td>• Name of outcome measure</td>
<td></td>
</tr>
<tr>
<td>• Unit of measurement</td>
<td></td>
</tr>
<tr>
<td>• Results</td>
<td></td>
</tr>
<tr>
<td>TREATMENT</td>
<td></td>
</tr>
<tr>
<td>BASELINE MEASURES:</td>
<td></td>
</tr>
<tr>
<td>END OF STUDY MEASURES:</td>
<td></td>
</tr>
<tr>
<td>FOLLOW-UP MEASURES:</td>
<td></td>
</tr>
<tr>
<td>Outcome assessor:</td>
<td></td>
</tr>
<tr>
<td>Blinded?:</td>
<td></td>
</tr>
<tr>
<td>Validated measurement tools used?:</td>
<td></td>
</tr>
<tr>
<td>CONTROL #1</td>
<td></td>
</tr>
<tr>
<td>BASELINE MEASURES:</td>
<td></td>
</tr>
<tr>
<td>END OF STUDY MEASURES:</td>
<td></td>
</tr>
<tr>
<td>FOLLOW-UP MEASURES:</td>
<td></td>
</tr>
<tr>
<td>Outcome assessor:</td>
<td></td>
</tr>
<tr>
<td>Blinded?:</td>
<td></td>
</tr>
<tr>
<td>Validated measurement tools used?:</td>
<td></td>
</tr>
<tr>
<td>CONTROL #2</td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>BASELINE MEASURES:</td>
<td></td>
</tr>
<tr>
<td>END OF STUDY MEASURES:</td>
<td></td>
</tr>
<tr>
<td>FOLLOW-UP MEASURES:</td>
<td></td>
</tr>
<tr>
<td>Outcome assessor:</td>
<td></td>
</tr>
<tr>
<td>Blinded?:</td>
<td></td>
</tr>
<tr>
<td>Validated measurement tools used?:</td>
<td></td>
</tr>
<tr>
<td>Items &amp; Guidelines</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Outcomes</strong> (continued)</td>
<td>□ Enough time for full follow-up of outcomes?:</td>
</tr>
<tr>
<td></td>
<td>□ Drop outs?:</td>
</tr>
<tr>
<td></td>
<td>□ Missing data?:</td>
</tr>
<tr>
<td></td>
<td>□ Proportion of participants followed-up:</td>
</tr>
<tr>
<td></td>
<td><strong>General comments:</strong></td>
</tr>
<tr>
<td><strong>Statistical Analysis</strong></td>
<td>□ Statistics used:</td>
</tr>
<tr>
<td></td>
<td>□ Does technique adjust for confounding?:</td>
</tr>
<tr>
<td></td>
<td>□ Attrition rate:</td>
</tr>
<tr>
<td></td>
<td>□ Was attrition adequately dealt with?:</td>
</tr>
<tr>
<td></td>
<td>□ Intention to treat:</td>
</tr>
<tr>
<td></td>
<td>□ Point estimates and measure of variability:</td>
</tr>
<tr>
<td></td>
<td><strong>General comments:</strong></td>
</tr>
<tr>
<td><strong>Results/Discussion</strong></td>
<td><em>(ADD A TABLE)</em></td>
</tr>
<tr>
<td></td>
<td><strong>General comments:</strong></td>
</tr>
<tr>
<td>Items &amp; Guidelines</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Funding or sponsorship</td>
<td></td>
</tr>
<tr>
<td>Reference list</td>
<td>☐ Searched for additional studies and reviews</td>
</tr>
</tbody>
</table>

Additional criteria for cohort studies:

☐ Are the groups assembled at similar time points in their disease progression?
☐ Were the groups comparable on confounding factors?
☐ Was there adjustment for the effects of these confounding variables?
☐ Was a dose-response relationship between intervention and outcome demonstrated?
☐ Were drop out rates and reasons similar between groups?

Additional criteria for case-control studies:

☐ Is the case definition explicit?
☐ Has the disease state of the cases been reliably assessed and validated?
☐ Were the controls randomly selected from the source of population of the cases?
☐ How comparable are the cases and controls with respect to potential confounding factors?
☐ Were interventions and other exposures assessed in the same way for cases and controls?
☐ Is it possible that over-matching has occurred in that cases and controls were matched on factors related to exposure?
## Appendix J

### Seeking, Understanding and Applying Evidence: Key Concepts and Findings

<table>
<thead>
<tr>
<th>Categories &amp; Definitions</th>
<th>Concepts &amp; Subcategories</th>
<th>Legal findings</th>
<th>Health policy findings</th>
<th>Health care findings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perspective of evidence</strong>&lt;br&gt;Population vs. individual perspective</td>
<td>- Population perspective&lt;br&gt;- Individual perspective</td>
<td>- Individualist perspective on evidence.&lt;br&gt;- Outcomes can have population impacts (e.g., class action cases).&lt;br&gt;- Findings of a trial judge can become a “public fact” (which can influence policy).&lt;br&gt;- Fact-driven, based on evidence placed before the judge for that particular case.</td>
<td>- Population perspective on evidence.&lt;br&gt;- Frequent portfolio changes increases reliance on others to examine issue.</td>
<td>- Combined population and individualist perspectives on evidence (population level research applied at individual level).</td>
</tr>
<tr>
<td><strong>Rules of evidence</strong>&lt;br&gt;A set of rules used to accept or reject evidence.</td>
<td>- Legal rules of evidence&lt;br&gt;- Health policy rules of evidence&lt;br&gt;- Health care rules of evidence</td>
<td>- Numerous formalized legal rules of evidence (e.g., admissibility of evidence).&lt;br&gt;- Judges fairly lenient in allowing lawyers to admit evidence and call experts (weight of evidence assigned later).&lt;br&gt;- Most judges not familiar with experts’ substantive area of expertise.&lt;br&gt;- Evidence is selectively disclosed.</td>
<td>- Structured review process for pharmaceutical evaluations.&lt;br&gt;- Recognition of hierarchy of evidence.&lt;br&gt;- Acceptable evidence dependant upon current political and bureaucratic culture.</td>
<td>- Interpretation of evidence based on practitioners’ educational training in research appraisal techniques.&lt;br&gt;- Appraisal skills maintained through continuing to read research and participation in formal review committees.&lt;br&gt;- Use of a mental template to review studies.&lt;br&gt;- Generalized knowledge of study design criteria.</td>
</tr>
<tr>
<td><strong>Process of change</strong>&lt;br&gt;Differing ways in which change occurs.</td>
<td>- Level of evidence required to implement change at the practitioner level.</td>
<td>No text references</td>
<td>No text references</td>
<td>- A few studies, from the right centres, utilizing the right methodologies.&lt;br&gt;- Endorsement of a technique by a “guru” in the field could change thinking and ideas, but not necessarily practice.</td>
</tr>
<tr>
<td>Categories &amp; Definitions</td>
<td>Concepts &amp; Subcategories</td>
<td>Legal findings</td>
<td>Health policy findings</td>
<td>Health care findings</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------</td>
<td>---------------</td>
<td>------------------------</td>
<td>---------------------</td>
</tr>
</tbody>
</table>
| Process of change (continued) | - Strategies for change  
- Court of Appeals  
- Common law | - Court of Appeals  
- Getting an issue onto the public and political agendas. | - Consult experts  
- Professional development  
- Reading literature  
- Involvement in research  
- Involvement in private practice  
- Presenting a novel idea to "shake things up" |
| | - Reasons for change  
- "Persuading" change | - To keep pace with changes in political parties.  
- To keep pace with changes in political priorities. | - To maintain academic appointment.  
- To maintain professional licensure  
- To adopt something they felt was "inspiring"  
- To find answers when a patient does not fit a textbook case.  
- "Influenced" to change |
| | - Barriers to change  
No text references | - Economic cycles | Private practice setting:  
- Limited exposure to outside influences of change.  
- Working long hours  
- Personal resistance to change  
- Reluctance to discredit alternative therapies. |
| | - Facilitators of change  
- Facts and findings that come out of failed litigation. | - Political factors  
- Economic factors  
- Litigation  
- Macro policy level: Stakeholders [public political agenda]  
- Micro policy level: Evidence | - Variety in practice setting: access to more experiences, research, professional development, time and money.  
- Personal motivation  
- Patients  
- Expert opinion  
- Media  
- Reluctance to discredit alternative therapies. |
<table>
<thead>
<tr>
<th>Categories &amp; Definitions</th>
<th>Concepts &amp; Subcategories</th>
<th>Legal findings</th>
<th>Health policy findings</th>
<th>Health care findings</th>
</tr>
</thead>
</table>
| **Adjudicators of evidence**<br>Differing uses of experts to adjudicate evidence. | - Experts used to provide opinion | - To persuade the judge what is medically appropriate and what is not.  
- Study author is the preferred expert witness (unless work is seminal, then others can report on it)  
- For literature to be admitted into evidence an expert must acknowledge that it is reputable and reliable.  
- Some lawyers pre-screen experts to determine whether they have provided evidence on similar cases, what they said, whether the judge liked it and whether they testify for one side or the other..  
- Can obtain transcripts and reports from lawyers in other cases.  
- "Go to" experts: those that frequently provide opinion. | - Commissioned researchers | - Peer reviewers |
| **Methods:**<br>- Review of expert's publications to find an article that could be used against their testimony.  
- Attacking the factual matrix of their opinion as opposed to their opinion itself (attacking an opinion is harder). If the facts are wrong, the opinion will not hold any weight.  
- "The art of advocacy is to try and make someone look biased". | No text references | - Peer reviewers |
<table>
<thead>
<tr>
<th>Categories &amp; Definitions</th>
<th>Concepts &amp; Subcategories</th>
<th>Legal findings</th>
<th>Health policy findings</th>
<th>Health care findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjudicators of evidence (continued)</td>
<td>- Experts used to conduct research</td>
<td>- Experts do not conduct research; they are commissioned to provide opinion on evidence at trial (their own research or their colleagues).</td>
<td>- Episodic activity - Experts commissioned to conduct policy research, if time permits.</td>
<td>- Inherent activity of the profession. - Experts conduct research on an ongoing basis.</td>
</tr>
<tr>
<td></td>
<td>- Experts used for peer review</td>
<td>No text references.</td>
<td>No text references.</td>
<td>- Ambivalence as to whether peer review is the best process. Reviews not conducted thoroughly. - Useful, constructive feedback. - Fosters improved research designs. - Some journals have reputations for being conservative.</td>
</tr>
<tr>
<td></td>
<td>- Court-appointed experts</td>
<td>- Rarely used in BC Courts - Role is to listen to evidence and assist judge with its interpretation. - Goes against the tenets of the adversarial process. - &quot;A way of purchasing bias without recognizing it as bias&quot;.</td>
<td>Not applicable</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Categories &amp; Definitions</td>
<td>Concepts &amp; Subcategories</td>
<td>Legal findings</td>
<td>Health policy findings</td>
<td>Health care findings</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Decision-makers of evidence</td>
<td>- Key players</td>
<td>Lawyers:</td>
<td>Stakeholders (the public political agenda):</td>
<td>Journal editors:</td>
</tr>
<tr>
<td>Who determines what evidence should be followed?</td>
<td>- Role: To persuade the judge to view the evidence from their perspective.</td>
<td>- Role: To make their issue a public one.</td>
<td>- Opinions divided on utility of journal editors.</td>
<td></td>
</tr>
<tr>
<td>Judges:</td>
<td>- Role: To determine validity of the evidence and to make a decision based this evidence lead by the lawyers.</td>
<td>- Target is higher macro-level policies.</td>
<td>- Journals get reputations for &quot;fair gatekeeping&quot; and being cautious</td>
<td></td>
</tr>
<tr>
<td>Not their role: To wade into the evidence and cross-examine witnesses.</td>
<td>- Evidence is only one part of the decision making process.</td>
<td>- Elected officials are the decision-makers.</td>
<td>- Cautious journals may be slower to publish studies with sensational outcomes.</td>
<td></td>
</tr>
<tr>
<td>- Decide weight of evidence and credibility of witnesses.</td>
<td>- Of the opinion that the Courts are &quot;crossing the line&quot; into the responsibilities of the legislature.</td>
<td>- May go against evidence.</td>
<td>- Cautious journals may encourage scientists to be cautious.</td>
<td></td>
</tr>
<tr>
<td>- Interpret existing law and its outcomes (legislature can change wording of the law)</td>
<td>- Courts need to become more robust in evaluating evidence.</td>
<td>- Evidence is only one part of the decision making process.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Can reject professional standards (e.g. standards of care).</td>
<td>- The adversarial system is not conducive to setting health policy.</td>
<td>- Courts of Appeal:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Courts of Appeal:</td>
<td>- Can over-turn lower Court rulings.</td>
<td>- Can over-turn lower Court rulings.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Can over-turn lower Court rulings.</td>
<td>Juries:</td>
<td>- Medical cases too technical for juries, therefore tried by judge alone.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juries:</td>
<td>- Affidavits, depositions, anecdotal, hearsay, standard of care, scientific evidence, legislation and precedent cases.</td>
<td>- Effectiveness, efficacy, cost effectiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Types of evidence</td>
<td>- Effectiveness, efficacy, cost effectiveness</td>
<td>- Hierarchy of evidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categories &amp; Definitions</td>
<td>Concepts &amp; Subcategories</td>
<td>Legal findings</td>
<td>Health policy findings</td>
<td>Health care findings</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Decision-makers of evidence (continued)</td>
<td>-Level of evidence</td>
<td>-Primary and secondary evidence (primary preferred)</td>
<td>-Secondary evidence</td>
<td>-Primary and secondary evidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Pharmanet and MSP data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Some &quot;institutional resistance&quot; to utilizing it.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Contracted researchers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Public opinion/input</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Depends on the policy issue</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Other forms of &quot;information&quot; (if evidence not available).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Barriers to decision making</td>
<td>-Judges decision making reliant on evidence presented by the lawyers.</td>
<td>-Multiple stakeholder groups each lobbying their issues.</td>
<td>No text references</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Judicial policy making limits policy options.</td>
<td>-Privacy issues around using Pharmanet and MSP databases.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Evidentiary basis for decision making</td>
<td>No text references</td>
<td>-Institutional resistance to using data.</td>
<td>No text references</td>
</tr>
<tr>
<td>Focus of application of evidence The focus of the application of the evidence</td>
<td>-Individual Population</td>
<td>Individual (client)</td>
<td>Population (health care system)</td>
<td>Individual (patient/client)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indirect: class of population</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix K
Interview Questions

Legal Professionals

Opening Questions
• How were you involved with Auton?
• Can you describe the process for you to obtain the required evidence to support a case?
• How much does a lawyer get to know about the scientific evidence?

Thematic Questions
1. Legal uses and interpretations of science-based medical evidence may diverge substantially from the uses and interpretation of that evidence by the medical and health care researchers who produce it and of the practitioners and health plans that use it in making clinical decisions and policies.
   • How do you decide what evidence is most important? (e.g., expert opinion, anecdotal evidence, research studies)
   • How do lawyers interpret medical evidence?
   • How is the evidence used?

2. Law, policy and health care differ in their conceptualization of the patient/client in either a population or individual perspective. Legal use of evidence has been described as "evidence of the instance" as it focuses on the tenets of individual rights, wrongs, and harms. Even though the law of evidence is a standard set of rules that overlooks particular individualized situations, law utilizes evidence in an effort to evaluate causation under particular circumstances.
   • Does the application of evidence ever have an outcome for the population as opposed to strictly for the individual? If so, provide an example.

3. Law, policy and health care differ in regard to the timing of utilization of evidence relative to the outcome. In law, evidence is used post-hoc to discern whether an outcome that has occurred in the past was caused by an appropriate, harmful or inappropriate action. It is used to judge responsibility and render justice.
   • Do you agree that in law evidence is only utilized in a post-hoc fashion? Explain.
   • In relation to Auton, do you recall how the role of evidence played out?

4. The way in which change occurs impacts each profession's conceptualization of evidence. In law, the change agent role is performed by either judges, tribunals or by a legislative action.
   • Who should perform the change agent role?
   • How would a dictum handed down by a higher Court impact subsequent proceedings?

5. Each profession utilizes a different approach to adjudicate differences. In the adversarial approach of the legal system, lawyers and judges perform the gatekeeper role by deciding which evidence can be presented to and admitted by the Court (e.g., voir dire).
6. Each profession subscribes to different rules of evidence.
   - How do you decide what evidence is admissible?
   - How do you determine standard of care?
   - How would you define effectiveness?
   - How would you define cost-effectiveness?
   - What 'level of evidence' do Courts use?

7. There are some differences between the professions in regard to the source of the evidence and who decides on the validity of the evidence (the "gatekeeper").
   - Who is the source of the evidence? (primary or secondary e.g., hearsay)
   - Who decides on the validity of the evidence (in a civil case)?

General Questions
1. Both lay and expert witnesses retained by each respective party can be naturally biased towards their attorney-employer's perspective.
   - What fields of expert witnesses are regarded more objective than others? Give examples.
   - How do you determine bias in expert witnesses' testimonies?
   - How do you look for additional evidence in a case?

2. Historically, admissibility of expert opinion has been based on an assessment of the expert's qualifications leaving an evaluation of the reliability and weight of their presented evidence up to the lawyers to discern.
   - How do you assess an expert witnesses' qualifications?
   - What about the use of Court-appointed experts?
   - How do you evaluate the reliability and weight of experts' evidence?
   - How do you evaluate the expert opinion on the merits of a study?

3. Does the adversarial structure of adjudication impede comprehensive information gathering.

4. What are some barriers to finding evidence?

5. What are some barriers to understanding evidence?

6. What are some barriers to using evidence?

7. What do you see as the major differences between the way evidence is used and interpreted by the three disciplines of law, health care and policy?

8. How could medical evidence be improved to interface better with the Courts?
Health Care Professionals

Thematic Questions

1. Legal uses and interpretations of science-based medical evidence may diverge substantially from the uses and interpretation of that evidence by the medical and health care researchers who produce it and of the practitioners and health plans that use it in making clinical decisions and policies.
   - How do you decide what scientific/medical evidence is most important?
   - How do you interpret medical evidence?
   - How is the evidence applied?

2. Law, policy and health care differ in their conceptualization of the patient/client in either a population or individual perspective. In response to the "evidenced-based medicine" movement, physicians are increasingly relying on population-level statistics based on probabilities in order to make clinical decisions; this equates to "evidence of the generalizable".
   - When you utilize evidence in your practice, do you use population-level evidence, such as randomized controlled trials, to make clinical decisions or do you view your patient in an individual perspective?

3. Law, policy and health care differ in regard to the timing of utilization of evidence relative to the outcome. In health care, evidence is generally used in a pre-hoc probabilistic fashion to support a clinical decision prior to a patient outcome. (However, medicine often reflects back on past decisions to learn from its mistakes (e.g., morbidity and mortality rounds).
   - Do you agree that in health care evidence is used in a pre-hoc fashion (e.g., before the clinical outcome) to support a clinical decision? Or, are there instances where evidence could be used after an outcome has occurred?
   - In relation to Auton, do you recall how the role of evidence played out?

4. The way in which change occurs impacts each profession's conceptualization of evidence. In health care, change may occur when new scientific evidence is adopted by key opinion leaders and eventually diffused to other practitioners.
   - How do changes in your clinical practice come into effect?
   - In your profession, who are the change agents?

5. Each profession utilizes a different approach to adjudicate differences. In health care, scientific evidence is generally accepted once it has been through a peer review process; editors become the adjudicator (or gatekeeper) of evidence for the profession.
   - What are your thoughts on the statement that the peer review process and publishers in particular, act as the gatekeeper of evidence for your profession?

6. Each profession subscribes to different rules of evidence.
   - How do physicians decide what evidence is acceptable?
   - How would physicians define effectiveness?
   - How would physicians define cost effective?
   - What 'level of evidence' do physicians use?

7. There are some differences between the professions in regard to the source of the evidence and who decides on the validity of the evidence (the "gatekeeper").
   - What source of evidence is utilized? (primary or secondary)
   - Are patients ever consulted on recommendations from EBM?
   - Who decides on the validity of the evidence?
General Questions

1. What are some barriers to finding evidence?
2. What are some barriers to understanding evidence?
3. What are some barriers to the use of evidence?
4. Does EBM improve clinical skills?
5. What are your thoughts on the validity of RCTs?
6. Do you feel medical evidence is evenly distributed amongst the medical specialties?
7. Does EBM control health care costs?
8. Does EBM improve health?
9. What do you see as the major differences between the way evidence is used and interpreted by the three disciplines of law, health care and policy?
Health Policy Professionals

Opening Questions

• When did you become involved with Auton?
• Can you describe the process for you to obtain the required evidence to make a policy decision?
• How much does the Minister get to know about the evidence?

Thematic Questions

1. Legal uses and interpretations of science-based medical evidence may diverge substantially from the uses and interpretation of that evidence by the medical and health care researchers who produce it and of the practitioners and health plans that use it in making clinical decisions and policies.

   - How do policy-makers decide what scientific/medical evidence is most important?
   - How do policy-makers interpret medical evidence?
   - How is the evidence used?

2. Law, policy and health care differ in their conceptualization of the patient/client in either a population or individual perspective. Within the policy forum, health care decision-makers are compelled to discern the fairest way of equitably distributing finite resources amongst a population of individuals. Within this population perspective, many competing needs must be weighed through various decision-making processes.

   - How do you prioritize a population's needs?
   - What about in the situation where you may have very strong evidence in support of an intervention that would benefit a few people vs. weaker evidence on an intervention that would benefit many?

3. Law, policy and health care differ in regard to the timing of utilization of evidence relative to the outcome. In policy, evidence is used pre-hoc to inform decision-makers on population-level issues prior to the outcome.

   - Do you use evidence differently when you are considering funding something new as opposed to when you are responding to a crisis?
   - In relation to Auton, do you recall how the role of evidence played out?

4. The way in which change occurs impacts each profession's conceptualization of evidence. In policy, change occurs in response to many political, legal and economic factors in an ad hoc fashion.

   - How does policy change in light of the political, legal and economic factors, ie: which is more powerful or important to affect change?
   - When would politics trump evidence? Can you give me an example?
   - When would evidence trump politics? Can you give me an example?
   - What about costs and equity? Can you give me an example?

5. Each profession utilizes a different approach to adjudicate differences.

   - How are differences between Ministries handled?
   - How are differences between bureaucrats and ministers handled?
   - What do policy-makers do if the evidence is non-existent?

6. Each profession subscribes to different rules of evidence.

   - How do policy-makers decide what evidence is acceptable?
   - What rules of evidence do policy-makers follow?
   - How do policy-makers define effectiveness?
   - How do policy-makers define cost-effectiveness?
• What are some barriers to finding evidence?
• What are some barriers to understanding evidence?
• What are some barriers to using evidence?
• Is the public ever consulted on policies?

7. There are some differences between the professions in regard to the source of the evidence and who decides on the validity of the evidence (the "gatekeeper").

• Who is the source of the evidence? (primary or secondary)
• Who decides on the validity of the evidence?
• Who is the decision-maker?

Judicial Review Questions
• What are your thoughts on the statement: "Judicial policymaking limits policy options"?
• What about: "The adversarial system impedes comprehensive information gathering"?
• And: "Judicial review favours national norms and standards"?

Additional Questions
• What role does evidence-based medicine play in regulating health care?
• What is your thought on the validity of RCTs?
• Does evidence control health care costs?
• Does evidence improve health?

Closing Questions
• What do you see as the major differences between the way evidence is used and interpreted by the three sectors of law, health care and health policy?
Appendix L
Scientific Evidence Pathways: Revised Framework

Legal Pathway

Perspective of Evidence

Rules of Evidence

Adjudicators of Evidence

Decision-Makers of Evidence

Process of Change

Focus of Application of Evidence

Health Care Pathway

Health Policy Pathway

GENERALIZABILITY OF OUTCOMES TO THE POPULATION (SOCIETY)

Beard, 2007
## Appendix M

#### Primary Data

<table>
<thead>
<tr>
<th>Study, Year</th>
<th>AHRQ* Exclusion Code</th>
<th>Description / Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eldevik, Eikeseth, Jahr &amp; Smith, 2006</td>
<td>NR-Design</td>
<td>Retrospective chart review</td>
</tr>
</tbody>
</table>

#### Secondary Data

<table>
<thead>
<tr>
<th>Study, Year</th>
<th>AHRQ* Exclusion Code</th>
<th>Description / Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson &amp; Romanczyk, 1999</td>
<td>NR-Review</td>
<td>Review of ABA programmatic and methodological elements; summary of behavioural research; debunking of ABA myths; ABA program models.</td>
</tr>
<tr>
<td>Bibby, Eikeseth, Martin, Mudford &amp; Reeves, 2001</td>
<td>NR-Design</td>
<td>No control group</td>
</tr>
<tr>
<td>Boyd &amp; Corley, 2001</td>
<td>NR-Design</td>
<td>Survey (case series)</td>
</tr>
<tr>
<td>Charman &amp; Howlin, 2003</td>
<td>NR-OCD</td>
<td>Conference proceedings. Description of methodological issues and recommendations for improving research.</td>
</tr>
<tr>
<td>Chorpita, Yim, Donkervoet, Arensdoerfer, Amundsen, McGrue, Serrano, Yates, Burns &amp; Morelli, 2002</td>
<td>NR-Design</td>
<td>Review of efficacy of treatments for disorders such as autistic disorder, anxiety disorders, depression.</td>
</tr>
<tr>
<td>Study, Year</td>
<td>AHRQ* Exclusion Code</td>
<td>Description / Reason for Exclusion</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Choutka, Doloughty &amp; Zirkel, 2004</td>
<td>NR-Review</td>
<td>Analysis of case law concerning ABA</td>
</tr>
<tr>
<td>Couper, 2004</td>
<td>NR-OCD</td>
<td>Opinion article on level of evidence to support EIBI funding.</td>
</tr>
<tr>
<td>Diggle, McConachie &amp; Randle, 2003</td>
<td>NR-Other</td>
<td>One of the two studies appraised in this systematic review did not use Lovaas; the second study (Smith, Groen and Wynn, 2000) is already an included primary study in this thesis.</td>
</tr>
<tr>
<td>Doughty, 2004</td>
<td>NR-Other</td>
<td>Not a rigorous systematic review (as per author's admission). Studies reviewed were not specific to Lovaas, but included other behavioural and skill-based programs. Any relevant studies were already included as primary studies in this thesis.</td>
</tr>
<tr>
<td>Drew, Baird, Baron-Cohen, Cox, Slaunwhite, Sweetenham, Berry &amp; Charman, 2002</td>
<td>NR-Design</td>
<td>Pilot RCT, however, &quot;parent training group&quot; intervention was not Lovaas.</td>
</tr>
<tr>
<td>Eikeseth, 2001</td>
<td>NR-OCD</td>
<td>Opinion article critiquing previous critical reviews on Lovaas.</td>
</tr>
<tr>
<td>Erba, 2000</td>
<td>NR-Review</td>
<td>Review of four early intervention programs: Discrete Trial Training (Lovaas); LEAP, floor time, and TEACCH.</td>
</tr>
<tr>
<td>Finch &amp; Raffaele, 2003</td>
<td>NR-Other</td>
<td>Six of the seven studies of this systematic review pre-date the literature search dates for this thesis. The seventh study (Smith, Groen &amp; Wynn, 2000) is already an included primary study in this thesis.</td>
</tr>
<tr>
<td>Fombonne, 2003</td>
<td>NR-Review</td>
<td>Review of epidemiological studies of pervasive developmental disorder.</td>
</tr>
<tr>
<td>Harris &amp; Delmolino, 2002</td>
<td>NR-Review</td>
<td>Review of the research on benefits of ABA.</td>
</tr>
<tr>
<td>Hayes, Inc., 2003</td>
<td>NR-Other</td>
<td>Excluded based on the cost of obtaining this proprietary information.</td>
</tr>
<tr>
<td>Howlin, 2003</td>
<td>NR-Review</td>
<td>Review of research on various autism interventions.</td>
</tr>
<tr>
<td>Hutchinson-Harris, 2003</td>
<td>NR-Design</td>
<td>Retrospective analysis of Lovaas treatment intensity</td>
</tr>
<tr>
<td>Study, Year</td>
<td>AHRQ* Exclusion Code</td>
<td>Description / Reason for Exclusion</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Kasari, 2002</td>
<td>NR-Review</td>
<td>Review of research studies on comprehensive interventions; suggestions for methodological improvements.</td>
</tr>
<tr>
<td>Lord, 2000</td>
<td>NR-OCD</td>
<td>Commentary on communication research in autism.</td>
</tr>
<tr>
<td>Ludwig &amp; Harstall, 2001</td>
<td>NR-Design</td>
<td>Critical review summary including: BCOHTA report (Bassett et al., 2000); Tregear, 2000 (included in this thesis); Smith, 1999 (reviews Lovaas studies which were previously critically appraised in the BCOHTA report (Lovaas, 1987 &amp; McEachin et al., 1993)).</td>
</tr>
<tr>
<td>Magiati &amp; Howlin, 2001</td>
<td>NR-Design</td>
<td>A report of a study conducted on use of two different cognitive assessment tools in autism.</td>
</tr>
<tr>
<td>Maley &amp; Mayberry, 2003</td>
<td>NR-Other</td>
<td>Abstract only. Contacted first author; study has not been published.</td>
</tr>
<tr>
<td>McGahan, 2001</td>
<td>NR-Design</td>
<td>Systematic review of the literature that summarizes evidence and expert opinions on behavioural therapy; not a critical appraisal.</td>
</tr>
<tr>
<td>Nelson &amp; Huefner, 2003</td>
<td>NR-OCD</td>
<td>Commentary on U.S. Individuals with Disabilities Education Act (IDEA) and its relation to current legal proceedings on access to Lovaas therapy.</td>
</tr>
<tr>
<td>Prior, 2004</td>
<td>ECL</td>
<td>Editorial on intensive behavioural intervention.</td>
</tr>
<tr>
<td>Rogers, 1998</td>
<td>NR-Review</td>
<td>Review of research; studies contrasted with established criteria for empirically supported treatments.</td>
</tr>
<tr>
<td>Sheinkopf &amp; Siegel, 1998</td>
<td>NR-Other</td>
<td>Previously critically appraised in BCOHTA report.</td>
</tr>
<tr>
<td>Smith, 1999</td>
<td>NR-Review</td>
<td>Systematic review of various treatments</td>
</tr>
<tr>
<td>Smith &amp; Lovaas, 1998</td>
<td>NR-OCD</td>
<td>Description of Lovaas' UCLA Young Autism Project. Previously critically appraised in BCOHTA report.</td>
</tr>
<tr>
<td>Smith, Buch &amp; Gamby, 2000</td>
<td>NR-Design</td>
<td>Small prospective study (N=6) on parent-directed intensive early intervention. However, no control group (multiple-baseline design instead - &quot;single-case experimental design&quot;).</td>
</tr>
<tr>
<td>Study, Year</td>
<td>AHRQ* Exclusion Code</td>
<td>Description / Reason for Exclusion</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Stahmer &amp; Ingersoll, 2004</td>
<td>NR-Design</td>
<td>Quasi-experimental study of an inclusive program (discrete trial training only a small part of program).</td>
</tr>
<tr>
<td>Tregear, 2000</td>
<td>NR-Other</td>
<td>Systematic review of Lovaas including three studies (Lovaas, 1987; McEachin et al., 1993; and Sheinkopf &amp; Seigel, 1998), which had been previously critically appraised in BCOHTA report.</td>
</tr>
<tr>
<td>Vastag, 2004</td>
<td>NR-OCD</td>
<td>Opinion article on UCLA’s Early Childhood Program.</td>
</tr>
<tr>
<td>Webster, Webster &amp; Feiler, 2002</td>
<td>NR-Review</td>
<td>Narrative review of controversies surrounding autism intervention; status of evidence; nature of research tools; methodological issues.</td>
</tr>
<tr>
<td>Weiss, 1999</td>
<td>NR-Design</td>
<td>Case series</td>
</tr>
<tr>
<td>Wolery &amp; Garfinkle, 2002</td>
<td>NR-Design</td>
<td>Systematic review of outcome measures used in intervention research in autism.</td>
</tr>
<tr>
<td>Zachor, Ethan &amp; Izhak, 2003</td>
<td>NR-Design</td>
<td>Conference poster presentation of program that mixes ABA with developmental approaches and regular preschool activities.</td>
</tr>
</tbody>
</table>

*A Agency for Healthcare Research and Quality (2002). Systems to rate the strength of scientific evidence.*
### Appendix N

#### Summary of Included Studies

<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen, Amerine-Dickens &amp;</td>
<td>Observational</td>
<td>N=42</td>
<td><strong>EARLY INTENSIVE BEHAVIOURAL TREATMENT</strong></td>
<td>Description: This research was an attempt to replicate the original Lovaas (1987) study within a community setting. The research site met all criteria set out by the National Institute of Mental Health for a Lovaas replication site.</td>
<td>INTAKE MEASURES: Groups were not similar on intake. The EIBT group had more ASD children, better-educated parents, and more dual parent families; these findings were statistically significant.</td>
</tr>
<tr>
<td>Smith, 2006</td>
<td></td>
<td>Ethnicity:</td>
<td>Based on Lovaas/UCLA manuals (1981 &amp; 2003)</td>
<td>Protocol in place to assess children every 6 months and to withdraw if not progressing.</td>
<td><strong>EARLY INTENSIVE BEHAVIOURAL TREATMENT (EIBT) GROUP:</strong> 1. IQ: statistically significant difference between groups with IQ increasing by 25 points (62 to 87) for EIBT group.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not reported</td>
<td>without aversives. Plus peer play training and &quot;advanced social skills&quot; training added after 2 years (not part of original Lovaas/UCLA model).</td>
<td><strong>Comparison group</strong></td>
<td>2. Merrill-Palmer (visual-spatial): Non-significant increase of 13 points.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment (EIBT)</td>
<td>Intensity: 35-40 hrs/week (&gt;3 yrs of age); 20-30 hrs/week (&lt;3 yrs of age).</td>
<td><strong>Recruitment:</strong> Referrals to the Central Valley Autism Project</td>
<td>4. Vineland (adaptive behaviour): Significant 9 point increase in Composite scores. Also, significant differences between groups in the Communication and Daily Living subscales along with a trend for Socialization.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n=21 (18 males)</td>
<td>Duration: 47 weeks/year x 3-5 years.</td>
<td><strong>Treatment assignment:</strong> &quot;...legal and ethical considerations precluded random assignment of children to groups&quot; (p. 5147). Therefore, a &quot;quasi-experimental&quot; design was used through a matching process. Parents had the option of choosing between various services outlined in the Matrix of Educational Options. If they had chosen EIBT, then a matching child was found by reviewing other children’s charts and choosing the first child who met the inclusion criteria and whose</td>
<td>5. Classroom placement: 17 of the 21 EIBT children were integrated into regular classrooms with varying assistance from &quot;shadow tutors&quot;.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AD: n=20</td>
<td>Parent training: 12-18 hours of training workshops. Participated in weekly training sessions. Encouraged to provide discrete trial training (not mandatory).</td>
<td>Parent had the option of choosing between various services outlined in the Matrix of Educational Options. If they had chosen EIBT, then a matching child was found by reviewing other children’s charts and choosing the first child who met the inclusion criteria and whose</td>
<td><strong>Inclusion criteria:</strong> 1) Diagnosis of autistic disorder (AD) or pervasive developmental disorder not otherwise specified (PDD-NOS). 2) IQ &gt;35 3) Chronological age 18-42 months at diagnosis and &lt;48 months at treatment onset. 4) No other medical condition</td>
</tr>
</tbody>
</table>
### Authors Study Design Sample Interventions Methods Results/Conclusions

<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen, Amerine-Dickens &amp; Smith, 2006 (continued)</td>
<td></td>
<td></td>
<td>Therapists/Training: 1. UCLA Consultants: Visited site 2-4 times per year x 3 years. 2. Supervisors: Graduate students with more than 2 years behavioural treatment experience. Completed internships at UCLA. Responsible for training the tutors. 3. Tutors: Community members. Main provider of direct services. Received 1-2 hours/month of training up to 1-2 years. 4. Site Director: One of the UCLA-trained supervisors. Responsible for overseeing each treatment plan.</td>
<td>parents had chosen another service.</td>
<td>1. BSID-R (IQ) 2. Merrill-Palmer (visual-spatial) 3. Reynell (language) 4. Vineland (adaptive behaviour) 5. Classroom placement Outcome assessors: Diagnosis made by an independent psychologist then confirmed by an examiner employed by the research site. All assessments were conducted by independent clinicians. However, blinding to group assignment could not be ensured.</td>
</tr>
</tbody>
</table>

**Dropouts**

- n=4 (EIBT)
- n=1 (comparison)

**“Enrolment period”**

1995-2000

**Therapists/Training:**

| 1. UCLA Consultants: Visited site 2-4 times per year x 3 years. |
| 2. Supervisors: Graduate students with more than 2 years behavioural treatment experience. Completed internships at UCLA. Responsible for training the tutors. |
| 3. Tutors: Community members. Main provider of direct services. Received 1-2 hours/month of training up to 1-2 years. |
| 4. Site Director: One of the UCLA-trained supervisors. Responsible for overseeing each treatment plan. |

**Ratio:** 1:1

**Comparison Group**

Eclectic community services selected by their parents from a variety of options, e.g.: *Special education classes*  *Early Start Autism Intervention Program*  *Home-based developmental intervention*  *Speech, occupational and behavioural therapy*

- The EIBT group did not show reliable IQ increases relative to the comparison group after the first year.
- The Lovaas/UCLA model can be implemented in a community setting.
- Increases in test scores were similar to Lovaas' results.
- The difference between the two treatment groups was smaller than in other studies because the comparison group also made gains.
- The addition of “Advanced Social Skills” training after the second year altered the replication of the Lovaas study. Therefore, only results from Years 1 and 2 would be directly comparable.
- Although the comparison group had gains in IQ, only one child advanced into a regular classroom with varying assistance from their shadow tutor.

**Authors' Conclusions:**

- The EIBT group did not show reliable IQ increases relative to the comparison group after the first year.
- The Lovaas/UCLA model can be implemented in a community setting.
- Increases in test scores were similar to Lovaas' results.
- The difference between the two treatment groups was smaller than in other studies because the comparison group also made gains.
- The addition of “Advanced Social Skills” training after the second year altered the replication of the Lovaas study. Therefore, only results from Years 1 and 2 would be directly comparable.
- Although the comparison group had gains in IQ, only one child advanced into a regular classroom with varying assistance from their shadow tutor.
### Study Design

<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen, Amerine-Dickens &amp; Smith, 2006 (continued)</td>
<td>Observational</td>
<td>N= 25</td>
<td>BEHAVIOURAL TREATMENT Based on Lovaas/UCLA manual (1981) without aversives.</td>
<td>Description: This study was designed to evaluate the outcomes achieved by 4-7 year old autistic children after receiving intensive behavioural treatment in a school setting for one year.</td>
<td>regular classroom. The authors hypothesized that the reason more EIBT children advanced may be due to the additional &quot;advanced social skills&quot; training that they received. There were significant differences between groups at intake.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ethnicity:</td>
<td>Intensity: Min 20 hrs/week (received 28 hours/week)</td>
<td>Duration: 1 year</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Norwegian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Behavioural treatment n=13 (8 males)</td>
<td>Parent training: Parents were trained as therapists in order to extend the therapy into the home. They worked alongside therapists in the school four hours/week x 3 months.</td>
<td>Recruitment: Participants were obtained from referrals to the research site. All referrals agreed to participate.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Eclectic treatment n=12 (11 males)</td>
<td>Therapists/Training: 1) Supervisors: Staff from the research site were responsible for setting up</td>
<td>Treatment assignment: An independent clinician assigned the children to either behavioural or eclectic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inclusion criteria: 1) Diagnosis of autism (per ICD-10 criteria); 2) Chronological age at intake between 4-7 years; 3) Deviation IQ of 50 or more, or ratio IQ of 50 or above; 4) Absence of other major medical conditions.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Therapists/Training: 1) Supervisors: Staff from the research site were responsible for setting up</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**INTAKE MEASURES:**

Although the groups were not significantly different on any of the intake variables, the eclectic group had higher average scores on 10 of the 11 variables resulting in the authors hypothesizing that the eclectic group could have been higher functioning. There were also more girls in the behavioural treatment group.

**BEHAVIOURAL TREATMENT GROUP:**

1. Wechsler (WPPS-I-R or WISC-R), or Bayley Scales. (Intellectual functioning): Significantly higher average scores than the eclectic group. 54% (n=7) achieved IQs >85.
2. Merrill-Palmer (visual-spatial): Higher average scores than the eclectic group (non-significant).
3. Reynell (language): Significantly higher average scores than the eclectic.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eikeseth, Smith, Jahr &amp; Eldevik, 2002 (continued)</td>
<td>Dropouts: 2 (1 in each group)</td>
<td>Enrolment period: November 1995 to November 1998.</td>
<td>the treatment programs. Supervisors had a minimum of 1,500 hrs of experience implementing the Lovaas UCLA program.</td>
<td>treatment, based on the availability of supervisors for behavioural treatment.</td>
<td>group. 31% (n=4) scores &gt;85. 4. Vineland (adaptive behaviour): Significantly fewer disruptive behaviours than the eclectic group. 15% (n=2) had scores &gt;85. Treatment goals: Similar to eclectic group in all areas with the exception of alternative/augmentative communication and behaviour management where the behavioural group was significantly less likely to have goals. Educational levels of therapists: Not significantly different between groups. Prediction of outcomes: Intake IQ, Intake Performance IQ and language were strong predictors of outcome measures. Intake Vineland predicted outcome Vineland and was positively correlated with other outcome measures (although they were non-significant). A significant correlation was found between intake age and change in Vineland Adaptive Behaviour scores (younger children made larger gains). In general, intake measures were not significant predictors of changes in scores. Although the behavioural group had higher average scores at follow-up, none of the individual measures were statistically significant.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2) Therapists were trained by Lovaas supervisors. Each child was assigned a special education teacher who provided 4-6 hours/week of treatment along with one or more aides, who provided the remaining treatment hours. Outside of treatment hours, children were assimilated into regular classrooms with shadow therapists. Therapists received 10 hours/week of supervision in addition to a weekly 2-hour meeting to revise and modify the child's treatment program.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ratio: 1:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ECLECTIC TREATMENT A variety of interventions (including Project TEACCH, sensory-motor therapies and Lovaas therapy). Interventions were designed to reflect &quot;best practices&quot; in autism treatment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Measurements: 1. # of hours per week of one-to-one treatment. 2. Treatment goals 3. Level of therapist training. 4. Wechsler (WPPSI-R or WISC-R), or Bayley Scales. (intellectual functioning) 5. Merrill-Palmer (visual-spatial) 6. Reynell (language) 7. Vineland (adaptive behaviour) 8. Prediction of outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Outcome assessors: Diagnosis made by an independent psychologist. All assessments were conducted by independent clinicians blinded to group assignment.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eikeseth, Smith, Jahr &amp; Eldevik, 2002</td>
<td></td>
<td></td>
<td>Setting: Within the public school system.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Intensity:</strong> Min 20 hrs/week (received 29.08).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Duration:</strong> 1 year.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Parent training:</strong> No parent participation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Therapists/Training:</strong> Each child was also assigned a special education teacher who provided 4-6 hours/week of treatment along with one or more aides, who provided the remaining treatment hours. Outside of treatment hours, children were assimilated into regular classrooms with shadow therapists. Therapists did not receive supervision, however they did receive a weekly 2-hour &quot;consultation&quot; with the supervisor and directors. <strong>Ratio:</strong> Variable (depending on program).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ECLECTIC TREATMENT GROUP:**

1. **Wechsler (WPPSI-R or WISC-R), or Bayley Scales. (intellectual functioning):** Lower average scores than the behavioural group. 17% (n=2) achieved IQs >85.

2. **Merrill-Palmer (visual-spatial):** Lower average scores than the behavioural group.

3. **Reynell (language):** Lower average scores than the behavioural group. 25% (n=3) had scores >85.

4. **Vineland (adaptive behaviour):** Significantly more disruptive behaviours than the behavioural group. 8% (n=1) had a score >85.

**Treatment goals:** Similar to the behavioural group in all areas with the exception of: alternative/augmentative communication and behaviour management where the behavioural group was significantly less likely to have goals.

**Educational levels of therapists:** Not significantly different between groups.

**Prediction of outcomes:** No correlations between intake age and outcome measures. Intake IQ was strongly associated with follow-up IQ and language, and positively but non-significantly correlated with follow-up Performance IQ and the Vineland in addition to correlating strongly with change in language.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eikeseth, Smith, Jahr &amp; Eldevik, 2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Authors' Conclusions:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Intensive behavioural treatment group made significantly larger improvements than the intensive eclectic intervention group, therefore, the specific type of treatment accounts for the gains as opposed to the intensity of treatment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Some 4-7 year old children may benefit from behavioural treatment as much as younger children.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Participants may have been higher functioning at intake.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• The study shows that behavioural treatment can be effectively adapted to the school setting for some children.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Intake IQ predicted outcome on many variables for the behavioural treatment group.</td>
</tr>
<tr>
<td>Authors</td>
<td>Study Design</td>
<td>Sample</td>
<td>Interventions</td>
<td>Methods</td>
<td>Results/Conclusions</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Eikeseth, Smith, Jahr &amp; Eldevik, in press</td>
<td>Observational</td>
<td>Same as Eikeseth et al., 2002</td>
<td>Same as Eikeseth et al., 2002</td>
<td>Same as Eikeseth et al., 2002. Plus:</td>
<td>INTAKE MEASURES: Same as Eikeseth et al., 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean age: 8 years, 2 months (both groups)</td>
<td>BEHAVIOURAL TREATMENT</td>
<td>ADDITIONAL MEASUREMENT:</td>
<td>BEHAVIOURAL TREATMENT GROUP: Five of the 13 children (38%) no longer required one-to-one treatment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean time between treatment onset and follow-up: 31.4 months</td>
<td>1. Achenbach Child Behaviour Checklist – Teacher Report Form (administered unblinded to both groups by the child's teacher).</td>
<td>1. IQ: statistically significant gains in IQ (25 points – from 62 to 87).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Received 28 hrs/week; reduced at age 6 to 18 hrs/week.</td>
<td></td>
<td>3. Achenbach Child Behaviour: No significant difference between groups on most scales with the exception of the behavioural group displaying fewer social problems and aggressive behaviour.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ECLECTIC TREATMENT</td>
<td>Educational levels of therapists:</td>
<td>No significant difference between groups. Teachers had 3 or more years of undergraduate or graduate training in special education or related discipline.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean time between treatment onset and follow-up: 33.3 months</td>
<td>Treatment goals: No significant difference between groups.</td>
<td>Scores within average range: 54% (7 of 13) children in the behavioural group scored within one standard deviation of the mean on IQ at follow-up. Also performed in the average range on most or all other measures.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Treatment hours:</td>
<td>Scores:</td>
<td>Prediction of outcomes: Age not associated with outcome or amount of change. Intake IQ strongly associated with follow-up IQ, language, and Vineland Scores (except Socialization). No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Received 29 hrs/week; reduced at age 6 to 16 hrs/week.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Study Design</td>
<td>Sample</td>
<td>Interventions</td>
<td>Methods</td>
<td>Results/Conclusions</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
<td>--------</td>
<td>---------------</td>
<td>---------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Eikeseth, Smith, Jahr &amp; Eldevik, in press (continued)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Correlation between intake IQ and gains in IQ, language, and adaptive scores.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>ECLECTIC TREATMENT GROUP:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>One of the 12 children (8%) no longer required one-to-one treatment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1. IQ: Non-significant gain of 7 points from 65 to 72.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2. Vineland (adaptive behaviour): Mean scores declined 6-12 points.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3. Achenbach Child Behaviour: No significant difference between groups in favour of eclectic treatment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Educational levels of therapists: No significant difference between groups. Teachers had 3 or more years of undergraduate or graduate training in special education or related discipline.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Treatment goals: No significant difference between groups.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Scores within average range: 17% (2 of 12) children in the eclectic group scored within one standard deviation of the mean on IQ at follow-up (an additional child did so both at intake and at follow-up). Exhibited clinically significant behaviour problems and scored below the average range on adaptive behaviour.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Prediction of outcomes: Intake Vineland scores were the strongest correlate with outcome scores (except Vineland Maladaptive). Intake IQ and intake language scores reliably associated with higher scores on...</td>
</tr>
<tr>
<td>Authors</td>
<td>Study Design</td>
<td>Sample</td>
<td>Interventions</td>
<td>Methods</td>
<td>Results/Conclusions</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>--------</td>
<td>---------------</td>
<td>---------</td>
<td>---------------------</td>
</tr>
</tbody>
</table>
| Eikeseth, Smith, Jahr & Eldevik, in press (continued) | | | | | follow-up IQ and language. Intake measures were more often associated with outcome scores and changes in scores for the eclectic group than for the behavioural group. Authors' Conclusions:  
- Behaviourally treated children had larger increases in scores than the eclectic group in addition to less abnormal behaviour and fewer social problems.  
- Most gains in IQ and communication in the behavioural group occurred between intake and Year 1.  
- Treatment may have to extend beyond one year in order to obtain gains in social behaviour and daily living skill.  
- Demographic variables and pre-treatment test scores did not predict outcomes.  
- Children who began behavioural therapy later (between 4 and 7 years) may still improve.  
- The only difference between the two groups between the 1 year follow-up and this study were in the Vineland Adaptive Behavioural Composite and the Vineland Socialization scores. |
<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Howard,</td>
<td>Observational</td>
<td>N=61</td>
<td>INTENSIVE BEHAVIOUR ANALYTIC INTERVENTION (IBT)</td>
<td>Description: Comparison of intensive behaviour analytic therapy to two controls: 1) Intensive</td>
<td></td>
</tr>
<tr>
<td>Sparkman, Cohen,</td>
<td></td>
<td></td>
<td>Setting: Home, school, community.</td>
<td>Eclectic; and 2) Generic early intervention (low intensity).</td>
<td></td>
</tr>
<tr>
<td>Green &amp; Stanislaw,</td>
<td></td>
<td></td>
<td>Intensity: 35-40 hrs/week (for &gt;3 years of age); 25-30 hrs/week (for &lt;3 years of age). (Actual: 25-40 hrs).</td>
<td>Study purpose: Not stated</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td>Duration: 14.21 months.</td>
<td>Recruitment: Referrals from regional centres.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parent training: Parents received training in basic behaviour analytic</td>
<td>Treatment assignment: Made by education officials, with parental input.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>strategies. Implemented program at home. Met with staff 1-2 times/month.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Therapists/Training: 4-5 instructional assistants (6-9 hrs/week/assistant).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Trained by Masters-level psychology or special education teachers.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Supervisors: Masters-level psychology or special education teachers.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ratio: 1:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>INTENSIVE ECLECTIC (AP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Setting: Special education classes.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Intensity: 25-30 hrs/week. (Actual: 30 hrs).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Duration: 13.25 months.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**INTAKE MEASURES:**
The authors reported that all three groups were similar at intake in regard to level of developmental delay. Mean scores on all measures were similar, except the nonverbal skills domain, which was significantly higher in the GP group. All groups had near normal learning rates for motor skills before intervention. Groups differed on age at diagnosis, intake, follow-up, and parental education.

**INTENSIVE BEHAVIOUR ANALYTIC (IBT):**
1. IQ: Thirteen children (49%) showed significant gains in IQ to >86 (3 children had normal IQs at intake but also showed significant increases).
2. Merrill-Palmer (visual-spatial): skills acquired at normal to above normal rates (statistically significant).
3. Reynell (language): Twenty children (69%) acquired expressive language skills at a normal to above normal rate (statistically significant).
4. Vineland (adaptive behaviour): skills acquired at normal to above normal rates (all statistically significant with the exception of the motor skills domain).

Children gained more than 14 months developmentally in nonverbal, receptive language, expressive language, and overall...
<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Howard, Sparkman, Cohen, Green &amp; Stanislaw, 2005 (continued)</td>
<td></td>
<td></td>
<td>Parent training: Not reported</td>
<td></td>
<td>communication, social, and motor skills.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Therapists/Training: Special Ed teacher and classroom aids.</td>
<td></td>
<td>INTENSIVE ECLECTIC (AP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Consultations with behavioural analysis grad students.</td>
<td></td>
<td>1. IQ: Only 2 children showed gains in IQ into the normal range (13%). (No children had normal IQs at intake).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ratio: 1:1 or 1:2 (therapist:child)</td>
<td></td>
<td>2. Merrill-Palmer (visual-spatial): skills acquired at close to normal rates.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NON-INTENSIVE EARLY INTERVENTION PROGRAM (GP)</td>
<td></td>
<td>3. Reynell (language): 1-2 children exhibited normal or above-normal acquisition of expressive language; while some declined.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Setting: Special Ed. Classes</td>
<td></td>
<td>4. Vineland (adaptive behaviour): Mean age equivalent gains in all domains were 7.53-12.63 months.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Intensity: 15 hrs/week. (Actual: 15 hrs).</td>
<td></td>
<td>NON-INTENSIVE EARLY INTERVENTION (GP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Duration: 14.75 months</td>
<td></td>
<td>1. IQ: Only 3 children showed gains in IQ (19%). Two children who had normal IQs at intake actually had lower IQs at follow-up.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parent training: Not reported.</td>
<td></td>
<td>2. Merrill-Palmer (visual-spatial): skills acquired at close to normal rates.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Therapist/Training: Special Ed teacher, or Speech Pathologist, and classroom aids.</td>
<td></td>
<td>3. Reynell (language): 1-2 children exhibited normal or above-normal acquisition of expressive language; while some declined.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ratio: 1:6 (therapist:child)</td>
<td></td>
<td>4. Vineland (adaptive behaviour): Mean age equivalent gains in all domains were 4.5-13.17 months.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Authors' Conclusions:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• No statistically significant differences between mean scores of the AP and GP groups.</td>
</tr>
<tr>
<td>Authors</td>
<td>Study Design</td>
<td>Sample</td>
<td>Interventions</td>
<td>Methods</td>
<td>Results/Conclusions</td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------------</td>
<td>--------</td>
<td>---------------</td>
<td>---------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Howard, Sparkman, Cohen, Green &amp; Stanislaw, 2005 (continued)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Children in IBT group outperformed AP and GP groups on virtually every follow-up measure.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Gains were smaller since children had only received 14 months of treatment.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Learning rates for IBT group matched or exceeded the normal rate of one year of development per year of age.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- 14 months is not enough for children to make up their differences compared with typically developing preschoolers.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Study dispels myth that high intensity of any intervention will bring about gains.</td>
</tr>
<tr>
<td>Authors</td>
<td>Study Design</td>
<td>Sample</td>
<td>Interventions</td>
<td>Methods</td>
<td>Results/Conclusions</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------</td>
<td>--------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Sallows &amp; Graupner, 2005</td>
<td>Randomized Controlled Trial</td>
<td>N=23</td>
<td>CLINIC-DIRECTED GROUP Lovaas autism treatment (UCLA Model) without aversives.</td>
<td>Description: Autistic children were matched on age and pre-treatment IQ then randomly assigned to groups.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ethnicity: Not reported</td>
<td>Setting: Clinic, then into mainstream preschool.</td>
<td>Study purpose: 1) To replicate the UCLA program in a community setting; 2) To determine whether residual symptoms of autism remain among children who achieve post-treatment test scores in the average range. 3) To determine whether any pre-treatment variables predict outcomes.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PRE-HOC GROUP ASSIGNMENT:</td>
<td>Intensity: Goal: 40 hrs/week. (Actual: 39 hrs/week (Year 1); 37 hrs/week (Year 2))</td>
<td>Recruitment: From local special education programs.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;Clinic-directed&quot; Group: n=13 (11 boys)</td>
<td>Duration: 4 years</td>
<td>Treatment assignment: Matched then randomized.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>REVISED POST-HOC GROUP ASSIGNMENT:</td>
<td>Therapists: &gt;18 years of age; 1 year of college.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;Rapid learners&quot;: n=11 (# boys: unknown)</td>
<td>Ratio: 1:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&quot;Moderate learners&quot;: n=12 (# boys: unknown)</td>
<td>Staff training: 30 hours of training. Senior therapists had college degrees, 1 year experience, plus a 16-week UCLA internship.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inclusion criteria:</td>
<td>PARENT-DIRECTED GROUP Lovaas autism treatment (UCLA Model) without aversives.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1) Chronological age between 24-42 months; 2) IQ ratio &gt;35; 3) neurologically normal; 4) Diagnosis of autism per DSM-IV.</td>
<td>Setting: Home, then into mainstream preschool.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Drop-outs: Two.</td>
<td>Intensity: 32 hrs/week (Year 1); 31 hrs/week (Year 2). # of hours per week were chosen by the parents.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>One child dropped out at one year; unknown group</td>
<td>Description: Autistic children were matched on age and pre-treatment IQ then randomly assigned to groups.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**INTAKE MEASURES**

The authors reported that the two groups were not equal at intake. Table 1 shows differences between the groups in relation to the number of nonverbal children (more in the clinic-directed group) and parental education. In addition, since the subgroup of "rapid learners" in the Clinic-Directed Group were found to have higher pre-treatment IQs, Vineland scores, and levels of verbal imitation, the researchers omitted these measures. Because the Parent-Directed Group ended up receiving a higher intensity of treatment than the researchers had expected, between group differences due to intensity were difficult to assess. Therefore, the children were identified as either "rapid learners" or "moderate learners" and then compared.

**RAPID LEARNERS**

1. IQ: 48% (n=11) Full Scale IQ increased from 55 to 104.
2. Significant gains in all areas measured, however, rate of increase over time, skill areas and children was not consistent.
4. Personality Inventory for Children: Scored in the average range on all scales except Factor III (they tended to worry).
5. Child Behaviour Checklist: Non-clinically significant range on all...
<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
</table>
| Sallows & Graupner, 2005 (continued) |                                                                              | assign; data not included. Second child switched to another provider of behavioural treatment after one year; data included. | Duration: 4 years
Parent Training: Weekly meetings (same as Clinic-Directed Group) Therapists: Same as Clinic Directed Group. Ratio: 1:1 Staff training: Same as Clinic-Directed Group. SUPPLEMENTAL TREATMENTS
• 22 of the children received supplemental treatment/services such as special education, speech therapy, sensory integration, diets, mega-vitamins. | Outcome assessors: Diagnosis made by independent child psychiatrists. Children identified as "rapid learners" (IQ >85 post treatment) were assessed by blinded psychologists. "Moderate learners" (no change in IQ) were assessed by the second author to reduce costs. | scales, except on Scale 3 (they worried more).
6. Vineland: Communication and Socialization ratings within average range.
7. ADI-R: Significant improvement on all three scales. Eight children scored in the non-autistic range and had their diagnoses removed by the child psychiatrist.
8. Residual Symptoms: Eight children increased their social skills to the adequate range. Three had borderline problems (all from the Parent-Directed Group).


MODERATE LEARNERS
1. IQ: Remaining 52% (n=12) did not show a significant increase in Full Scale IQ.
2. Some increases in developmental age equivalents.
4. Personality Inventory for Children: More tantrums, difficulty interacting with others, and more learning problems.
5. Child Behaviour Checklist: Less interactive, more pre-occupied, less attentive, and more easily frustrated.
7. ADI-R: No significant differences.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Design</th>
<th>Sample</th>
<th>Interventions</th>
<th>Methods</th>
<th>Results/Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sallows &amp; Graupner, 2005 (continued)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>COMBINING ALL CHILDREN</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Average Full Scale IQ gain from 51 to 76.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Significant gains were found in Full Scale IQ, Verbal IQ, Performance IQ, receptive language, Vineland Communication, Vineland Socialization, ADI-R Social skills, and Communication.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Authors' Conclusions:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Ability to imitate is highly correlated with Full Scale IQ, Language, and Social Skills.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Children with higher pre-treatment IQs more likely to reach IQs in the average range in 4 years.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Post treatment IQ best predicted by Early Learning Measure, pre-treatment IQ, ADI-R (impairment in social interaction), and ADI-R (Communication) scores.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Social skill and language acquisition predicted by pre-treatment ability to imitate.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Number of weekly treatment hours less related to outcome.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- No correlation between # of hours of parental involvement in treatment and outcomes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- # of hours of peer play was significantly related to ratings of social skills.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- No correlation between supplemental therapies and outcomes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- UCLA EIBT program can be implemented in a community.</td>
</tr>
<tr>
<td>Authors</td>
<td>Study Design</td>
<td>Sample</td>
<td>Interventions</td>
<td>Methods</td>
<td>Results/Conclusions</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------</td>
<td>--------</td>
<td>---------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Sallows &amp; Graupner, 2005 (continued)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- UCLA program can be replicated without use of aversives.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Parent-directed children did about as well as clinic-directed children.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- High intensity of treatment and intensive supervision do not compensate for low levels of pre-treatment skills.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- IQ&lt;44 and an absence of language predicted limited progress.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Rate of learning, imitation, and social relatedness predicted favourable outcomes.</td>
</tr>
<tr>
<td>Authors</td>
<td>Study Design</td>
<td>Sample</td>
<td>Interventions</td>
<td>Methods</td>
<td>Results/Conclusions</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Smith, Groen &amp; Wynn, 2000</td>
<td>Randomized Controlled Trial</td>
<td>N=28 (autism: n=14; PDD-NOS: n=14)</td>
<td>INTENSIVE TREATMENT: Based on Lovaas autism treatment manual (1981) plus normal classes. Setting: Home x 1 year then normal classroom. Intensity: 30 hours/week (5 hrs provided by parents). (Children only received 24.52 hrs/week during their first year). Classroom hours not reported. Duration: 2-3 years (Mean treatment duration was 33.44 months). Parent training: Not reported for this group. Therapists: 4-6 student therapists (supervised). Parents provided 5 hrs/week. Ratio: 1:1 x 1 year then small groups. Staff training: Supervisors trained in Lovaas.</td>
<td>Description: Children with autism and PDD-NOS were matched and then randomized to receive either intensive treatment based on Lovaas, or &quot;Parent Training&quot; (also based on Lovaas). Study purpose: To extend the literature on the effectiveness of early intervention with other developmental disorders besides autism (e.g.; PDD-NOS) and to address some of the issues in regard to cost (by designing a less intensive intervention) and the stress on children and families. Recruitment: Referrals to UCLA Young Autism Project. Treatment assignment: Matched pair (diagnosis &amp; IQ), random assignment. Measurements: 1. Stanford Binet or Bayley Scales of Infant Development (IQ) 2. Merrill-Palmer (visual-spatial) 3. Reynell (language) 4. Vineland (adaptive) 5. Early Learning Measure 6. Achenbach &amp; Teacher Form (follow-up only), 7. Weschler (academic) (follow-up only) 8. Class placement</td>
<td>INTAKE MEASURES: No significant differences between groups at intake. INTENSIVE TREATMENT GROUP: 1) IQ: Statistically higher scores than Parent-Training. 2) Merrill-Palmer (visual-spatial): Statistically higher scores than Parent-Training. 3) Reynell (language): Statistically higher scores than Parent-Training. 4) Vineland (adaptive behaviour): No difference between groups. 5) Classroom Placement: Less restrictive school placements. 6) Weschler (academic): Higher scores than Parent-Training. Two children in the Intensive Treatment group met criteria for &quot;best outcome&quot; (placement in regular classes without special services and IQ&gt;85). Two additional children met the school placement criteria, but had an IQ below the cut-off. One of these four children continued to display significant behavioural problems. PARENT TRAINING GROUP: 1) IQ: Statistically lower scores than the Intensive Treatment group. 2) Merrill-Palmer (visual-spatial): Statistically lower scores than the Intensive Treatment group. 3) Reynell (language): Statistically lower scores than...</td>
</tr>
<tr>
<td>Authors</td>
<td>Study Design</td>
<td>Sample</td>
<td>Interventions</td>
<td>Methods</td>
<td>Results/Conclusions</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------</td>
<td>--------</td>
<td>---------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Smith, Groen</td>
<td></td>
<td></td>
<td>Intensity: Unclear.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&amp; Wynn, 2000</td>
<td></td>
<td></td>
<td>Appears to have been 10 hours/week (5 hrs training</td>
<td>9. Family satisfaction</td>
<td>the Intensive Treatment group.</td>
</tr>
<tr>
<td>(continued)</td>
<td></td>
<td></td>
<td>sessions + 5 hours parental sessions) for first 3-9</td>
<td>Outcome assessors: Diagnosis made by independent psychologists.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>months. Then, decreasing to 5 hours/week of</td>
<td>Outcome assessments conducted by blinded doctoral psychology students</td>
<td>No difference between groups.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>parental sessions only thereafter.</td>
<td>(except for the Early Learning Measure (ELM) and the Achenbach</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Duration: Not reported.</td>
<td>which were administered in an unblinded fashion. However, the researchers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>attempted to compensate by testing the ELM for inter-rater reliability</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parent training: 5 hours/week x 3-9 months.</td>
<td>utilizing another group of blinded assessors.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Meetings with first author every 3 months.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Therapists: Parents</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ratio: 1:1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Staff training: N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Note: The first four patients entered into the</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>study were initially exposed to contingent aversives, however this practice was later stopped.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Authors' Conclusions:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Some children with PDD-NOS will benefit from treatment as much as the autistic children.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Intensive treatment may be more effective than parent training.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Between-group differences in follow-up IQ were only half that reported by McEachin et al. (1993) (16 vs. 31 points), as were the # of children placed in regular classes without further services (27% vs. 47%).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Results did not support McEachin's findings that intensive treatment reduced behavioural problems and improved adaptive functioning.</td>
</tr>
</tbody>
</table>
### Appendix 0
Analysis of Heterogeneity Between Groups

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical Heterogeneity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groups relatively equal in size?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Equal sex ratio?</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Equal distribution of diagnoses?</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>✓</td>
<td>N/A</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Equal distribution of chronological age at intake?</td>
<td>NR</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Equal distribution of age at diagnosis?</td>
<td>✓</td>
<td>NR</td>
<td>NR</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Similar parental levels of education?</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Similar parental marital status?</td>
<td>X</td>
<td>NR</td>
<td>NR</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Similar IQ at intake?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Similar rates of imitation at intake?</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>X</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Similar drop out/lost to follow-up rates?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Groups similar at intake?</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Interventions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment fidelity met?</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Treatment duration goal met?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Non-Lovaas control group?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Standardized Lovaas treatment protocols for experimental group?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Standardized therapist training for experimental group?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CRITERIA</td>
<td>STUDY</td>
<td>Early Intensive Behavioural Therapy (EIBT)</td>
<td>Behavioural treatment</td>
<td>Intensive Behavioural Analytic Treatment (IBT)</td>
<td>Early Intensive Behavioural Treatment</td>
<td>Intensive treatment</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------------------</td>
<td>-------------------------------------------</td>
<td>-----------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>Name of experimental treatment</td>
<td>Cohen, Amerine-Dickens &amp; Smith, 2006</td>
<td>Early Intensive Behavioural Therapy (EIBT)</td>
<td>Behavioural treatment</td>
<td>Intensive Behavioural Analytic Treatment (IBT)</td>
<td>Early Intensive Behavioural Treatment</td>
<td>Intensive treatment</td>
<td></td>
</tr>
<tr>
<td>Regular classroom inclusion?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Shadow tutor?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not reported</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Setting(s)</td>
<td>Home</td>
<td>Public schools</td>
<td>Public schools</td>
<td>Home</td>
<td>Clinic (Clinic-directed group) Home (Parent-directed)</td>
<td>Home</td>
<td></td>
</tr>
<tr>
<td>Therapist/child ratio</td>
<td>1:1 (at home)</td>
<td>1:1 (at school)</td>
<td>1:1 (at home)</td>
<td>1:1 (initially at home)</td>
<td>1:1 (at home)</td>
<td>1:1 (at home)</td>
<td></td>
</tr>
<tr>
<td>Peer play training</td>
<td>Yes</td>
<td>15-60 min sessions, 3-5 times/week</td>
<td>Yes</td>
<td>Unknown intensity</td>
<td>Yes</td>
<td>Unknown intensity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>Unknown intensity</td>
<td>Yes</td>
<td>Unknown intensity</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>------------------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Intensity (per protocol)</td>
<td>35-40 hrs/wk (&gt;3 yrs)</td>
<td>20 hrs/wk</td>
<td>20 hrs/wk</td>
<td>35-40 hrs/wk (&gt;3 yrs)</td>
<td>25-30 hrs/wk (&lt;3 yrs)</td>
<td>Clinic: 40 hrs/wk</td>
<td>30 hrs/wk</td>
</tr>
<tr>
<td></td>
<td>20-30 hrs/wk (&lt;3 yrs)</td>
<td>Reduced to 0</td>
<td>20 hrs/wk</td>
<td>Reduced to 0</td>
<td>20 hrs/wk</td>
<td>Home: decided by parents (M=32 hrs/wk)</td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>3-5 years</td>
<td>1 year</td>
<td>31.4 months (Chronologic age of 8 months)</td>
<td>14 months (interim data)</td>
<td>4 years</td>
<td>2-3 years</td>
<td></td>
</tr>
<tr>
<td>Teaching format</td>
<td>Discrete trial training</td>
<td>Discrete trial training</td>
<td>Discrete trial training</td>
<td>Discrete trial training</td>
<td>Discrete trial training: and, &quot;Procedures supported by subsequent research&quot; (generic)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Therapists/Tutors</td>
<td>Recruited from the community.</td>
<td>Special Ed teacher &amp; Aids</td>
<td>Special Ed teacher &amp; Aids</td>
<td>College students</td>
<td>1st yr college students</td>
<td>Students</td>
<td></td>
</tr>
<tr>
<td>Supervisors/senior therapists</td>
<td>Grad students with UCLA internships</td>
<td>Trained in UCLA Lovaas model with 1500 hrs experience</td>
<td>Trained in UCLA Lovaas model with 1500 hrs experience</td>
<td>Masters-prepared generic ABA-trained</td>
<td>Grad students with UCLA internships</td>
<td>Student therapists trained in UCLA Lovaas model with 1500 hrs experience.</td>
<td></td>
</tr>
<tr>
<td>CRITERIA</td>
<td>STUDY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of aversives?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Additional non-behavioural treatments?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes (n=22) special ed, speech therapy, sensory integration.</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents as formal therapists?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (in &quot;Parent-Directed Group&quot;).</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measures include tests of intelligence, non-verbal skills, visual-spatial, language, adaptive skills, maladaptive behaviour?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Un-validated tools</td>
<td>Reynell</td>
<td>Reynell</td>
<td>Achenbach</td>
<td>Reynell</td>
<td>Reynell Achenbach</td>
<td>Reynell</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>------------------------------------</td>
<td>-----------------------------------------</td>
<td>------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td></td>
<td>Methodological Heterogeneity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Randomization?</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Similar outcome measure time points?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Independent diagnosis?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Blinded outcome assessment?</td>
<td>P</td>
<td>✓</td>
<td>X</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Concealed allocation?</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Quality rating</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Poor</td>
<td>Moderate</td>
<td>High*</td>
</tr>
<tr>
<td></td>
<td>Heterogeneity in Results</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IQ</td>
<td>SS</td>
<td>SS</td>
<td>SS</td>
<td>SS</td>
<td>NSS</td>
<td>SS</td>
</tr>
<tr>
<td></td>
<td>Merrill-Palmer</td>
<td>NSS</td>
<td>NSS</td>
<td>NR</td>
<td>SS</td>
<td>NSS</td>
<td>SS</td>
</tr>
<tr>
<td></td>
<td>Reynell:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comprehension</td>
<td>TTS</td>
<td>SS</td>
<td>NR</td>
<td>SS</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td></td>
<td>Expressive</td>
<td>NSS</td>
<td>SS</td>
<td>NR</td>
<td>SS</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>NR</td>
<td>SS</td>
<td>NR</td>
<td>NR</td>
<td>NSS</td>
<td>SS</td>
</tr>
<tr>
<td></td>
<td>Vineland adaptive behaviour:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Composite</td>
<td>SS</td>
<td>SS</td>
<td>SS</td>
<td>SS</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td></td>
<td>Communication</td>
<td>SS</td>
<td>SS</td>
<td>SS</td>
<td>SS</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td></td>
<td>Daily Living</td>
<td>SS</td>
<td>SS</td>
<td>SS</td>
<td>SS</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td></td>
<td>Socialization</td>
<td>TTS</td>
<td>NSS</td>
<td>SS</td>
<td>SS</td>
<td>NSS</td>
<td>NSS</td>
</tr>
<tr>
<td></td>
<td>Motor</td>
<td>NR</td>
<td>SS</td>
<td>NR</td>
<td>SS</td>
<td>SS</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Vineland maladaptive behaviour</td>
<td>N/A</td>
<td>N/A</td>
<td>SS</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------</td>
<td>------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>--------------------------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td>Regular classroom placement rates**:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group: (Total)</td>
<td>81%</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>UTA</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>- Without support</td>
<td>28%</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>UTA</td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>- Fading out tutor</td>
<td>19%</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>NR</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>- Full support</td>
<td>33%</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>NR</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Control group 1 (Total)</td>
<td>5%</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>NR</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>- Without support</td>
<td>NR</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>NR</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>- With support</td>
<td>NR</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>NR</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Achenbach (Teacher Report Form):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Withdrawed</td>
<td>N/A</td>
<td>NSS</td>
<td>NSS</td>
<td>N/A</td>
<td>UTA</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Somatic</td>
<td>N/A</td>
<td>NSS</td>
<td>NSS</td>
<td>N/A</td>
<td>UTA</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Anxious/Depressed</td>
<td>N/A</td>
<td>NSS</td>
<td>NSS</td>
<td>N/A</td>
<td>UTA</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>N/A</td>
<td>NSS</td>
<td>NSS</td>
<td>N/A</td>
<td>UTA</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Thought</td>
<td>N/A</td>
<td>NSS</td>
<td>NSS</td>
<td>N/A</td>
<td>UTA</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Attention</td>
<td>N/A</td>
<td>NSS</td>
<td>NSS</td>
<td>N/A</td>
<td>UTA</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Delinquent</td>
<td>N/A</td>
<td>NSS</td>
<td>NSS</td>
<td>N/A</td>
<td>UTA</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Aggressive</td>
<td>N/A</td>
<td>NSS</td>
<td>NSS</td>
<td>N/A</td>
<td>UTA</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Learning rates:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-verbal</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Receptive</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Expressive</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Self-help</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NSS: Not Sufficiently Supported; SS: Sufficiently Supported; UTA: Unreported Treatment Analysis; NR: Not Reported**
<table>
<thead>
<tr>
<th>CRITERIA</th>
<th>STUDY</th>
<th>Woodcock-Johnson III</th>
<th>Early Learning Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cohen, Amerine-Dickens &amp; Smith, 2006</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Elkeseth, Smith, Jahr &amp; Eldevik, 2002</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Elkeseth, Smith, Jahr &amp; Eldevik, in press</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Howard, Sparkman, Cohen, Green &amp; Stanislaw, 2005</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Sallows &amp; Graupner, 2005</td>
<td>UTA</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Smith, Groen &amp; Wynn, 2000</td>
<td>CS</td>
<td>CS</td>
</tr>
</tbody>
</table>

**Key:**
- NR: Not reported
- N/A: Not applicable
- SS: Statistically significant differences
- NSS: Not statistically significant
- TTS: Trend towards significance

**P:** Partial
**UTA:** Unable to analyze as per a priori design due to confounding.
**CS:** Clinically significant

*Based on initial quality assessment scores. Subsequent analysis determined this study to be of moderate quality.

**Full integration into school and “shadow tutor” support.**
### Appendix P

#### Direction of Biases

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Heterogeneity</td>
<td>Group size</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>↑</td>
<td>↑★</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Sex ratio</td>
<td>0</td>
<td>↓</td>
<td>↓</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Diagnoses</td>
<td>↓★</td>
<td>N/A</td>
<td>N/A</td>
<td>0</td>
<td>N/A</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Chronological age</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
<td>↑★</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Age at diagnosis</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>↑★</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Parental levels of education</td>
<td>↑★</td>
<td>N/A</td>
<td>N/A</td>
<td>↑★</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Parental marital status</td>
<td>?</td>
<td>N/A</td>
<td>N/A</td>
<td>?</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>IQ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Rates of imitation</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>?</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Drop out/lost to follow-up rates</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Interventions</td>
<td>Treatment fidelity</td>
<td>⇆</td>
<td>↑</td>
<td>↑</td>
<td>⇆</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Treatment duration goal</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Non-Lovaas control group</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Standardized Lovaas treatment</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>protocols for experimental group</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Standardized therapist training</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>↓</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
## Appendix P

### Direction of Biases

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity</td>
<td>NR</td>
<td>O</td>
<td>O</td>
<td>NR</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Teaching format</td>
<td>O</td>
<td>O</td>
<td>C</td>
<td>C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Additional non-behavioural</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>treatments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of un-validated tools</td>
<td>⇨</td>
<td>⇨</td>
<td>⇨</td>
<td>⇨</td>
<td>⇨</td>
<td>⇨</td>
</tr>
<tr>
<td>Methodological Heterogeneity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomization?</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>0</td>
</tr>
<tr>
<td>Similar outcome measure time</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>points?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent diagnosis?</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Blinded outcome assessment?</td>
<td>↑</td>
<td>0</td>
<td>0</td>
<td>↑</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Concealed allocation?</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**KEY:**
- **O** No reported between group differences
- **↑** Bias in direction of experimental group
- **↓** Bias in direction of control/comparison group
- **⇔** Neutral
- **?** Unknown
- ★ Statistically controlled
- **N/A** Not applicable
- **NR** Not reported
- **C** Potential confounding