SELECTING QUALITY INDICATORS FOR PULMONARY REHABILITATION PROGRAMS IN CANADA: A MODIFIED RAND APPROPRIATENESS METHOD STUDY

by

Walden A. Cheung

B.Kin., The University of British Columbia, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Rehabilitation Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2017

Abstract

Pulmonary rehabilitation (PR) is a comprehensive intervention of self-management education and exercise training that improves quality of life, exercise tolerance, symptoms of dyspnea, and reduces the risk of hospitalization in patients living with chronic respiratory diseases such as chronic obstructive pulmonary disease, asthma, lung cancer, and interstitial lung disease. Despite the proven benefit of pulmonary rehabilitation, recent studies have found notable inconsistencies in its organization and delivery. Inconsistencies within clinical practice are likely to affect the quality in the delivery of pulmonary rehabilitation. Quality indicators (QIs) are tools similar to a checklist that can potentially remediate these concerns. While other jurisdictions have created quality indicators for pulmonary rehabilitation programs, their methodological approach to developing these quality indicators is questionable. This study developed 56 quality indicators with a rigorous approach using a modified RAND Appropriateness Method. A panel comprising twelve PR healthcare professionals and stakeholders was created to create a list of QIs. The panel rated each indicator based on four criteria (importance, scientific soundness, reliability, and feasibility) and listed which indicator they believed could determine a quality pulmonary rehabilitation program. This study recommends that the 56 QIs, based upon consensus, be used for operationalizing the evaluation and auditing of PR programs as well as for establishing clinical benchmarks.

Lay Summary

Pulmonary rehabilitation (PR) is a comprehensive intervention of education, self-management, and exercise that improves quality of life, exercise tolerance, and symptoms of breathlessness in patients living with chronic respiratory diseases. However, despite the proven benefit of this intervention, there remains a challenge in translating PR from the research laboratory to clinical practice. A recent multinational survey found notable variation in the organization and delivery of PR programs, including variation in the organization of healthcare professionals, setting of PR, reporting of outcomes, and use of outcome measures. Such organizational variation is concerning as it occurs in areas of PR that are linked to the quality and effectiveness of the programs. The goal of this study was to develop a tool to assess the quality of PR programs. This study developed quality indicators to potentially identify specific areas of a PR program that need improvement.

Preface

This thesis contains the work of the candidate, Walden A. Cheung, under the supervision of Dr. Pat G. Camp. Study design was a collaborative effort between Dr. Pat G. Camp and Walden A. Cheung. The candidate, Walden A. Cheung, was primarily responsible for data collection, analysis, and interpretation.

This study presented in this thesis received ethical approval from the Providence Health Care Research Ethics Board (UBC-PHC REB Number: H16-01678). All data were collected at the Pulmonary Rehabilitation Research Laboratory at St. Paul's Hospital, Vancouver, British Columbia.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

COPD: Chronic Obstructive Pulmonary Disease

HRQoL: Health Related Quality of Life

USA: United States of America

UK: United Kingdom

PR: Pulmonary Rehabilitation

PALs: Physical Activity Levels

ADLs: Activities of Daily Living

mMRC: Modified Medical Research Council

ILD: Interstitial Lung Disease

AACVPR: American Association of Cardiac and Pulmonary Rehabilitation

ATS: American Thoracic Society

ERS: European Respiratory Society

CTS: Canadian Thoracic Society

CIHI: Canadian Institute for Health Information

KT: Knowledge Translation

QIs: Quality Indicators

RAM: Rand/UCLA Appropriateness Method

MRAM: Modified RAND Appropriateness Method

DI: Disagreement Index

MET: Metabolic Equivalent

SEPAR: Sociedad Española de Neumología y Cirugía Torácica

BTS: British Thoracic Society

Acknowledgements

This thesis is dedicated to the many people who provided me with incredible support.

To my thesis committee Dr. Pat G. Camp, Dr. Chris Ryerson and Dr. Andrea Gershon: thank you for your support, guidance, and critique throughout my project, which have improved my work tremendously. I would like to extend my gratitude to the members of the Pulmonary Rehabilitation Research Laboratory: Abbie Wright MSc, Ori Ben-Ari (MSc Candidate), Ali Alasmari (PhD Student), Jessica Inskip PhD, and Ashley Kirkam BSc, for their assistance in this study.

To all the volunteers who gave up their summers to help me: Stella, Michelle, Tina and Gloria. I am tremendously grateful for your time, help, and support. To Ellis and Ann: thank you for allowing me to work part-time as an exercise therapist throughout my degree and thereby to gain an invaluable amount of clinical experience. To Sion, Kate, Kris, George, Andrew, and Gabriel: thank you for the support and for consistently providing the best coffee brews I've ever had. And finally, to my parents, my brother Josh, and the friends I love the most, (Jacqueline, Mia, Hilury, Eva, Minhee, Julie, and Ethan): thank you for reading all my short stories, and for all the support and love you provided.

Chapter 1: Background and Rationale

1.1 Pulmonary Rehabilitation

Pulmonary rehabilitation (PR) is acknowledged as one of the cornerstone therapies for individuals suffering from Chronic Obstructive Pulmonary Disease (COPD) and other chronic lung diseases[1]. Why is this so? Perhaps because the treatment of COPD requires more than the use of pharmaceuticals[2]. COPD is a pulmonary disease defined as "*persistent airflow limitation that is usually progressive and associated with an enhanced chronic inflammatory response in the airways and lung to noxious particles and gas*[3]. It is a disease also confounded with many co-morbidities, such as depression and anxiety, cardiovascular disease, and extra-pulmonary problems including exercise intolerance, peripheral muscular dysfunction and very low physical activity levels (PALs). Such a complex etiology, understandably, requires a complex intervention that treats the disease from multiple dimensions[1]. Pharmacological therapy is a standard and crucial treatment to relieve the airflow limitation and dyspnea COPD patients suffer from as well as improve their quality of life and exercise tolerance[1]. However, such therapies do not necessarily treat the extra-pulmonary problems of COPD, such as leg fatigue, psychological distress and muscle atrophy, which are proven to increase the risk of premature mortality and hospitalization[1,2]. The addition of PR to pharmacological therapy helps a patient improve both the pulmonary and non-pulmonary complications from which they are suffering [2]. This approach likely extends to other chronic lung diseases such as interstitial lung disease (ILD), lung cancer, asthma, and cystic fibrosis[1].

There is strong scientific evidence for the benefit of PR for COPD and other chronic lung diseases. Perhaps more importantly, PR has been shown to improve patient outcomes outside the

research laboratory and in clinical practice[4]. Yet challenges still remain in PR delivery and implementation, and particularly in its quality assurance[5,6]. As PR programs expand away from hospital-related settings and into settings with less resources, scientists and clinicians must ensure that the quality of the program delivery remains. In this introduction, we briefly review the benefits and components of PR, the challenges of quality assurance, and an approach to resolve some of these challenges.

PR is defined by the American Thoracic Society/European Respiratory Society (ATS/ERS)[1] as a "*comprehensive intervention based on a thorough patient assessment followed by patient-tailored therapies that include, but are not limited to, exercise training, education, and behaviour change, designed to improve the physical and psychological condition of people with chronic respiratory disease and to promote the long-term adherence to health-enhancing behaviours*." PR is an integration of exercise training, education and self-management training that aims to work with standard pharmaceutical treatment to improve health outcomes such as dyspnea, quality of life and exercise tolerance.

1.1.1 Benefits of PR

There are many symptoms a COPD patient experiences, including dyspnea, fatigue, anxiety, and depression. These symptoms can severely impact the patient's ability to work, their engagement in physical activity, and their overall quality of life, ultimately leading to increased rates of hospitalization, health-care utilization, and early mortality[1]. The most recent meta-analysis and systematic review regarding PR found that overall, PR conferred clinically-meaningful improvements to a COPD patient's health-related quality of life (HRQoL)[7]. It was also

established that clinically-meaningful improvements occurred in exercise tolerance and functional status[7]. Consequently, the Cochrane Airways editorial board[8] decided to close further systematic reviews on the efficacy of PR in COPD, stating that there is enough evidence to show that PR improves "*patient-oriented outcomes*" in COPD, specifically in HRQoL and exercise-tolerance domains. They also cited recent evidence showing that PR can decrease the rates of hospitalization and premature mortality in COPD[1,6].

The scientific evidence supporting the efficacy and benefit of PR is substantial, but what about the clinical evidence? A recent audit on PR programs in England and Wales surveyed over 200 PR programs with 7000 COPD participants[4]. The audit found that for every 100 patients that participated in a PR program, 83 participants had an improvement in exercise capacity and functional status. Of the 83 participants, 63 had clinically-meaningful improvement. Similar numbers were found for health status and HRQoL assessments, where 74 participants had improvement, of which 61 had clinically-meaningful improvement. Overall, both the scientific literature and clinical data show that COPD patients that participate in a PR program have an increased HRQoL, exercise tolerance, and decreased number of hospitalizations[1,4,7].

There is also evidence showing that PR, and specifically exercise training, can improve HRQoL, exercise intolerance, dyspnea, and functional status in chronic respiratory diseases other than COPD; these include asthma, interstitial lung disease (ILD), non-cystic fibrosis bronchiectasis, lung transplantation and lung cancer[9–13].

1.1.2 Components of PR: Exercise Training

This section introduces the main components of PR: exercise training and self-management education. Exercise training is the foundation of PR. Of the two components of PR, exercise training is the most important[1]. Exercise training is defined as "*planned, structured, and repetitive bodily movements (that) are performed to improve or maintain one or more components of physical fitness*[14]." Exercise training follows a structure based on frequency, intensity, type of training, and time. In order to induce a physiological adaptation, each component of the structure must be followed and progressed[15]. The purpose of exercise training in PR is not to train the lungs but to train the cardiovascular and musculoskeletal system. It is thought that improvements of these systems are the mechanism that improves the symptoms of dyspnea, as well as exercise tolerance and HRQoL. Thus, the purpose of exercise training in PR is to improve exercise tolerance and reduce symptoms of dyspnea and fatigue so that COPD patients can improve their functional capacity and have the ability to perform activities of daily living (ADLs) that they may not have been able to perform otherwise. Together, a combination of resistance training and endurance training are found to provide the most benefit as they are able create physiological improvements while resolving some of the exercise limitations of COPD patients.

The ATS/ERS[1] PR statement recommends that both endurance and resistance training be used in PR. The exercise prescription should be tailored to fit participants' needs. This is because co-morbidities and subjective experiences of dyspnea differ between participants. The improvement of peripheral musculature conditioning via exercise training may lead to reductions of dynamic hyperinflation thereby reducing aspects of dyspnea[1]. Improvements in dyspnea are significant

because of the association of dyspnea with exercise tolerance, mortality, and PALs[1]. Other benefits exercise training may confer are improvements in psychological outcomes such as anxiety and depression, cardiovascular function, and HRQoL[1].

1.1.2.1 Endurance Exercise Training in PR

The conventional exercise training interventions in PR are endurance exercise training and resistance training. Endurance exercise training is defined as exercise training that "*involves large muscle groups in dynamic activities that result in substantial increases in heart rate and energy expenditure…resulting in improvements in the function of the cardiovascular system and the skeletal muscles, leading to an increase in endurance performance*[14]." Improvements in cardiovascular and musculoskeletal systems can lead to a participant performing ADLs at a higher exertion without being limited by fatigue and dyspnea. There is a wide variety of endurance exercises for PR participants including stationary cycling and treadmill walking. Endurance exercise training follows several exercise guidelines which will be discussed in the *Delivery of PR Programs* section (see section 1.2 below).

1.1.2.2 Resistance Training in PR

The other exercise training intervention used in PR is resistance training[1]. Resistance training is defined as exercise training "*designed specifically to increase muscular strength, power, and endurance by varying the resistance, the number of times the resistance is moved in a single group (set) of exercise, the number of sets done, and the rest interval provided between sets*[14]." The purpose of resistance training is to improve peripheral muscle function for fall prevention, increase bone mineral density, and enhance functional exercise capacity[1]. Resistance training

focuses on improving the muscle function of both the upper and lower limbs. This is especially practical for the participant, as many ADLs, such as dressing, washing dishes, brushing teeth, and combing hair, involve the upper limbs. A systematic review by O'Shea *et al.*[17] found that resistance training may improve the relevant musculature and allow the participant to perform more ADLs without experiencing as much fatigue and dyspnea. Similar to endurance exercise training, resistance training follows several exercise guidelines which we will discuss in the *Delivery of PR Programs* section (see section 1.2 below).

1.1.2.3 Alternative Exercise Training

In addition to endurance and resistance exercise training, there are alternative or adjunct forms of exercise training that have been used in PR, such as neuromuscular electrical stimulation, inspiratory muscle training, whole body vibration, neuromuscular electrical stimulation, and Tai Chi[18–20]. While active research is being undertaken around the efficacies of these alternate interventions[18-20], the ATS/ERS[1] states that the evidence for benefit is inconsistent and as a result these alternative forms of exercise training are not currently recommended by PR guidelines.

1.1.3 Components of PR: Education and Self-Management

The other component of PR is education for the purposes of self-management. This component has been integrated with exercise training to promote behavioural changes such as adherence to exercise, smoking cessation and managing symptoms[1,21]. The overall purpose is to "*promote self-efficacy (i.e., the confidence in successfully managing one's health) through increasing the patient's knowledge and skills required to participate with health care professionals in optimally managing their illness.*[1]" For example, in PR, COPD patients may be educated first about their

disease, their symptoms and how to use their medication or supplemental oxygen. This knowledge is then integrated with behavioural strategies that help the patient manage their symptoms. Participants work together with the PR staff in creating an action plan for adverse exacerbation events, maintaining an active lifestyle, breathing techniques, energy conservation techniques, smoking cessation techniques, and nutritional changes[22]. Bourbeau *et al.*[22] reported that self-management in COPD patients reduces utilization of healthcare resources and leads to improvement in HRQoL. The implementation of self-management education in PR has the ability to improve self-efficacy which, as mentioned before, is related to behavioural change and longer adherence to exercise after discharge from PR[22].

1.1.4 Components of PR: Maintenance

Unfortunately, the benefits of PR appear to wane six to twelve months after discharge from the program. The reasons for this vary and include progression of disease and decline of physical activity after discharge[1]. The loss of benefits after discharge can once again put the patient at risk for hospitalizations and pre-mature mortality. However, a recent study by Spencer *et al.*[23] reported that a maintenance program was able to keep the improvements of exercise capacity and HRQoL in participants 12 months after discharge from PR. Thus, it is important that PR programs provide strategies (follow-up, home-exercise prescription, or referral to community wellness programs) to maintain the exercise training, self-management, and physical activity levels attained during the PR program.

1.2 Delivery of Pulmonary Rehabilitation

The focus of PR research has now moved to expanding the delivery of its service to diseases other than COPD and settings other than in-hospital settings[1,24]. One of the challenges facing of the expansion of PR programs to non-traditional settings (with the possibility of comparatively fewer resources), is quality-control and reducing the amount variation in clinical practice[6,25]. Below we will discuss the recommended organization and delivery of a PR program.

1.2.1.1 Organization and Delivery of PR

A recent policy statement by the ATS/ERS[6] describes the organization and standards a qualified PR program must have. While different countries have different health authorities and therefore different organizational structures and outcome measures in their PR programs, the PR programs still need to remain grounded on scientific literature[25,26]. A PR program must have: 1) a structured exercise training protocol that is supervised; 2) an education and self-management component to teach participants, promote physical activity and evaluate outcome measures; and 3) strategies for maintenance of exercise and physical activity after discharge[6]. Additional components of PR may include psychological and nutritional counselling. The following sections will detail the specific components of a PR program.

1.2.1.1.1 Duration and Frequency of PR Program

The ATS/ERS[6] recommends that PR programs to be at least 8 weeks in duration to confer benefits for participants. Participants should attend the programs 2-3 times a week, with each session being 1-4 hours long[6]. Sessions should be broken into parts for exercise training and self-management education to take place.

1.2.1.1.2 Staffing

There are many healthcare professionals involved in delivering PR. The following staffing organization is recommended by Jenkins *et al.*[27]. The essential staff involved must be able cover the three core aspects of PR. This staff thus must consist of 1) a respirologist; and 2) a physical therapist, respiratory therapist, or exercise specialist. Auxiliary staff include 1) a nurse or 2) an occupational therapist or 3) a pharmacist or 4) a dietician or 5) a social worker or 6) a psychologist. The respirologist usually oversees the initial clinical evaluations, medication management, supplemental oxygen prescription, and referral to the PR program. This staff member needs to be available to mediate any medical-related problems that may arise during the program. Next, a physical therapist, respiratory therapist, or certified exercise specialist is responsible for exercise testing and training. The education and self-management components can involve both the essential and auxiliary staff. For example, the respiratory therapists and physical therapists may be involved with educating the PR participants with airway clearance techniques and energy conservation strategies. A respiratory nurse may be responsible for the self-management portion of the PR program. Finally, the pharmacist may focus education specifically on medication use. The ATS/ERS reports no that there is no evidence-based support for a certain PR staff-to-patient ratio but emphasizes that the staff must be proficient and comfortable enough with the number and type of participants they are supervising during exercise training and education[1]. For example, the ATS/ERS states that North American jurisdictions use 1:8 for educational sessions, 1:4 for exercise training, and 1:1 for complex PR participants, while the UK jurisdiction uses 1:8 for exercise training and 1:16 for educational sessions[1,5]. It should be noted that due to certification and continued education, some staff roles can extend to different areas of PR. For example, a respiratory therapist may receive an

American College of Sports Medicine exercise certification to supervise exercise training in a PR program. Other auxiliary staff can provide additional components to a PR program. For example, the psychologist is responsible for evaluating psychological conditions and treating the depression and anxiety common in participants with chronic lung diseases. The nutritionist provides dietary evaluation and support. The social worker provides various health services to the participant, while the occupational therapist evaluates and modifies the participant's home to improve its accessibility. Together the staff functions to help the participant transfer the skills they have learned from the program into their daily life in order to adhere to exercise and preserve the benefits of PR after program discharge.

1.2.1.1.3 Safety Equipment and Emergency Response

Emergency Response

The setting of PR programs can vary from outpatient hospital settings, to community settings, physical therapy clinics, homes and tele-rehabilitation programs in rural settings[5,28]. In order for all PR clinics to be safe and effective for the participants, safety and emergency equipment must be present[6]. At a minimum, emergency response protocols must be in place for PR programs in all different settings. For example, PR programs away from the hospital-setting should have access to an ambulance, a defibrillator, and a first-aid kit with specific emergency medications[27].

Safety Equipment

Safety must also be a priority during exercise assessment and training of PR participants. Jenkins *et al.*[27] state that the following safety equipment is necessary for a PR program: a pulse oximeter, heart rate monitor, sphygomomanometer, portable oxygen tank and nasal prongs. Participants

must be pre-screened for any co-morbidities or resting abnormalities that could affect their health during an exercise assessment or session. Monitoring of a patient's oxygen saturation, symptoms, and heart rate must continue throughout exercise training as well. As all eligible chronic lung diseases for PR have varying degrees of response to exercise it is expected that healthcare professionals are educated in safety.

Exercise Equipment

Jenkins *et al.*[27] recommend that the following exercise equipment be in place for PR: hand weights (for resistance exercise assessment/training), stairs or a step (for endurance exercise assessment/training), a walking track (for endurance exercise assessment/training), and a stopwatch (for exercise assessment/training).

1.2.1.1.4 Patient Assessment

The ATS/ERS[6] recommends that PR program staff perform a thorough assessment of the eligible participant's health status, functional exercise capacity, muscular strength/endurance, dyspnea, and psychological symptoms. Having these measures will allow exercise training to be safe, precise and effective. As there is variation between eligible disease populations and individual participants, staff must also evaluate a patient's individual goals and needs[1,27].

1.2.1.1.5 Exercise Training

While the types of exercise training in PR may differ in modality and purpose, effective exercise training follows several exercise principles, such as a prescribed frequency, intensity, time, and

type[5]. Exercise training must also be progressed to overload the skeletal muscle system to confer benefit[5].

Endurance Exercise: Continuous Endurance Training

The prescribed frequency for endurance exercise training should be 3-5 days a week and the majority of sessions should range from 20-60 minutes[16,29]. The prescribed level of intensity should be above 60% of a participant's maximal work rate using an objective measure[16,29]. This maximal work rate is prescribed based on the initial exercise assessment and may be modified using subjective measures, such as the Borg Scale, during exercise[16]. The initial exercise assessment is crucial for exercise training as it helps determine an individual's maximal work rate, and notifies the staff of any safety issues to take into consideration for a participant. The type of exercise training is determined by what resources are available in the PR program, such as walking, stationary cycling, or treadmill training.

Endurance Exercise: Interval Training

Should endurance training cause dyspnea or fatigue and disrupt the training session, a different form of endurance exercise has been suggested[29]. Interval training is a modification of endurance exercise where a prolonged exercise session is broken into several breaks, but requires the participant to exercise at higher intensity (80-100% of maximal work rate)[29]. Interval training leads to less disruptions in exercise and lower scores of symptoms for participants while conferring the same benefits of exercise tolerance and HRQoL[30]. This form of endurance training may be particularly useful for participants whose disease is more severe.

Resistance Training

The ATS/ERS[1] reports there is no optimum prescription for resistance training for individuals with chronic lung disease and instead follows the recommendation of the American College of Sports Medicine (ACSM). The ACSM[29] recommends that the frequency of resistance training should be 2-3 days a week, with 8-12 repetitions, and 1-3 sets. The intensity of resistance training should be 60-70% of a participant's one repetition maximum (an objective measure) which is performed during the initial assessment[29].

1.2.1.1.6 Education and Self-Management

The topics of education and self-management include: smoking cessation, use of medication, energy conservation techniques, airway clearance techniques, nutrition, the importance of exercise, and supplemental oxygen[1]. This area of PR should be covered by a trained staff member who is able to provide the recommended education and behavioral-changing strategies to improve a participant's health outcomes[27].

1.2.1.1.7 Outcome Measures in PR

To properly evaluate the effect of PR program on a participant, key outcome measures must be evaluated before and after completion of the program. The ATS/ERS[6]. statement includes the following as essential and standard outcome measures: functional exercise capacity, health status, dyspnea, and psychological symptoms Feedback from a participant's experience and economic measures, such as health-care utilization, are also recommended[6].

With regards to outcome measures the degrees of change should follow those of scientific

evidence, meaning that outcome measures of participants should be grouped according to their disease and benchmarked to minimal clinically-important differences in the scientific literature and compared with other programs[27]. The minimal clinically-important difference (MCID) is the "*smallest difference in a measurable clinical parameter that indicates a meaningful change in the condition for better or for worse, as perceived by the patient, clinician or investigator*[1]." This will allow stakeholders involved in PR to grasp the impact PR has on individuals with chronic lung diseases.

Evaluating these outcome measures with the proper scales and questionnaires is important as they follow scientific evidence. The following assessments must be made before and after a PR program: physical measurements (height, weight, age, spirometry), exercise capacity (field-based walking test or cardiopulmonary exercise test), peripheral muscle strength of upper and lower limbs, physical activity, HRQoL, symptoms and psychological status (anxiety and depression), self-efficacy, and health-care utilization. Additional measures may be required and/or modified for certain diseases. For example, composite measures which measure body mass index, airway obstruction, dyspnea, and exercise capacity (BODE Index[31]) or measure activity, dyspnea, and airway obstruction (ADO[31]) and balance outcome measures may be needed for participants with COPD while MCID values are different for those with ILD, or pulmonary hypertension[1]. It is recommended that audits of PR programs occur annually and include the following components: what outcome measures are taken, patterns of attendance and frequencies of referral, patient feedback, and how the processes of PR are implemented[27].

1.2.2 PR Programs in Non-Hospital Settings

With an increasing demand for PR programs in more rural and community areas, several studies have investigated whether or not PR in non-hospital settings, such as tele-rehabilitation settings, remote-delivery via videoconferencing, or home-based programs confers similar benefits to PR delivered in a traditional hospital setting[32,33]. This is particularly pertinent for the rural population in Canada, who have a greater prevalence of chronic respiratory diseases such as COPD, ILD and who also tend to have greater mortality and hospitalization rates[28]. The distance and remoteness of rural settings leads to more limitations such as travel and cost for eligible PR participants. With increasing technology being utilized in tele-healthcare services, the delivery of PR through these programs offers advantages of accessibility and costs for many eligible participants who cannot participate in more centralized hospital-based PR programs.

In comparison to face-to-face delivery of PR in a hospital setting, remote-delivery PR programs are performed through live-videoconferencing where a physical therapist leads, for example, five different PR programs[28,33]. Each different program has ongoing supervision from a health care professional or rehabilitation assistant and uses remote-pulse oximetry for safety monitoring[28]. Tele-monitoring technology allows participants and their health-care providers to monitor participants' health-related outcomes in the PR program. A recent review by Goodridge and Marciniuk[28] reported several studies that found that these remote-delivery programs conferred similar improvement in exercise tolerance and HRQoL to hospital-based PR programs, and that the use of remote pulse oximetry was feasible. While the expansion of PR programs into community, rural, and remote settings is enabling more eligible participants to access PR, there is the concern that these programs do not have the same resources as hospital-based PR

programs. This includes the number and type of staff, and the type of exercise equipment. The differences in resources could lead to variations in the delivery of PR programs.

1.3 Issues of Unwarranted Variation in PR

Variation is found in many healthcare services[34]. It is important to distinguish necessary variation from unwarranted variation[34]. Some variation across healthcare services is necessary to be effective in treatment, since diseases are heterogeneous, and variation is present between patients[35]. Nevertheless, there are many procedures in healthcare service where variation should be low, such as the sanitary practices of hand-washing and wearing protective clothing[34,35].. Unwarranted variation may negatively impact the quality of the healthcare service leading to inefficiencies and medical errors which both have consequences of increased costs and mortality. Therefore, for many stakeholders and patients using healthcare services, the lack of quality control and consistency of practice is cause for concern.

1.3.1 Evidence of Unwarranted Variation in PR

Currently, one of the healthcare services that may have unwarranted variation is PR[5,24]. Understandably, as PR is used to rehabilitate a heterogeneous patient population, there is necessary variation in the delivery of the service[25]. Such variation ensures that each participant receives care tailored to their goals and needs[25,36]. However, akin to hand-washing, there are still elements of PR that should not be varied as they are fundamental to the quality and success for every PR participant[25]. Recently, several studies have described this variation[4,24,37,38].

A recent study by Spruit *et al.*[5] surveyed 430 PR programs from 40 countries around the world and found that while most PR programs (83.4%) considered QoL as a very important health outcome, other crucial patient-centered outcomes in PR were not considered as important. Below are a few examples. In North America, only 35.8% of PR programs considered dyspnea as an important outcome. Only 9.1% of PR programs in North America, and 3.2% in Europe considered depression as an important outcome[5]. Approximately one third of PR programs from both North America and Europe considered physical activity (21.9% and 33.5%) and activities of daily living (36.4% and 28.7%) as important[5]. Healthcare utilization was not considered as an important outcome in the surveyed programs. Most significantly, this study found that issues of safety, patient satisfaction, accessibility, and efficiency were not considered as the most important outcomes in PR. Interestingly, 17 PR programs were found to have only one staff member[5].

A study by Yohannes *et al.*[37] found similar results to that of Spruit *et al.*[5]. Of the 239 PR programs audited in the United Kingdom, 51% of PR programs did not fully meet the standards put forth by the audits. Only 33% of programs provided maintenance programs. 47% of programs only partially met the standard of having a PR-trained healthcare professional supervising participants in exercise training[37]. Fifteen programs did not have any staff supervising at all. Only 148 programs (69% of programs) fully measured the recommended health outcomes (health status, exercise) for program participants, while 60 (28%) programs partially measured these outcomes. Six (7%) of programs did not have a medical director or coordinator for the program[37].

A report by Steiner *et al.*[4] looked at 210 programs in 2015 with 7000 participants and found the following: 31% of referred participants did not undergo an initial assessment and 40% of those who completed an initial assessment did not complete the program. Additionally, of the 210 programs, five programs offered no endurance exercise, a crucial component of exercise training. Eighteen percent of programs did not provide an exercise prescription or they had a non-standard way of prescribing exercise, while 52% used a 'best-guess' estimate of exercise[39]. While many programs provided resistance training (94% of programs), most of those programs (70% of the 94%) provided a resistance training prescription through subjective measures. The outcome measures recorded in many of the programs were also low, as muscle strength was only recorded in 22% of programs, physical activity in 34% of programs, and 35% of programs did not provide participants with written exercise plans[4].

The findings of this study were similar to a very recent study by Dechman *et al*[38]. Of 112 programs they audited in Canada, over 100 programs (90%) delivered resistance training, but only 39 of them properly used an exercise test to prescribe a training intensity[38]. The authors found that the most common method of prescribing exercise intensity was through oxygen saturation and dyspnea, both of which do not follow the recommended guidelines of exercise prescription[16,29,38]. Another paper by Camp *et al.*[24] using the same dataset found that 11 (9%) programs had only one PR staff member, 39 programs (30%) of programs did not evaluate participant outcomes after program completion, and 68 programs (53%) did not provide any maintenance sessions. Furthermore, 26 programs did not have emergency equipment or protocol in place and 13 programs (10%) did not have supplemental oxygen for exercise training[24].

These studies carry a consistent theme: there is notable variation in the delivery of PR services around the world[5]. The studies suggest that there are inconsistencies in the implementation of PR programs which may affect the quality of PR outside the research lab[5,6]. The heterogeneity found in many PR programs may be in part due to different methods of funding and different health authority policies, as well as the variety of chronic respiratory disorders affecting patients apart from COPD, the different stages of disease severity, and the alternate settings in which PR programs are implemented[1]. However, there is still the expectation that PR programs follow a protocol that ensures quality and consistency as per scientific recommendations[6]. The purpose of consistency is to identify key program elements that may improve or worsen desired outcomes, therefore allowing accurate changes to occur. What are the possible problems that could arise when PR programs deviate from the recommendations? These will be discussed below.

1.4 Defining Quality for Evaluating Healthcare Services

Before the consequences of deviating from quality are discussed, we must define quality. Quality is a broad term that can carry very different meanings depending the context of use[40]. Defining and measuring the quality of a healthcare service requires defining the dimensions that can provide a context for quality[40]. According to the Canadian Institute for Health Information (CIHI)[41], the properties of a quality Canadian healthcare service can be defined by four dimensions:

1. Access to comprehensive, integrated healthcare service: The delivery of the healthcare meets the expectations of the participants' desired setting and time.

2. Quality, Appropriateness, and Safety of healthcare service: The delivery of the healthcare service provides a safe and evidence-based intervention.

3. Efficiency and effectiveness of healthcare service: The delivery of the healthcare service provides an intervention that is beneficial to a participant's health outcomes and ideal for its own resources.

4. Patient experience with healthcare service: The delivery of the healthcare service is able to tailor to the participant's individual preferences.

These four dimensions will provide the context to measure the quality of healthcare service performance in this study. As each dimension has an amount of variability in the context of evaluating quality, these dimensions can be considered variables in the performance of a healthcare service. If a healthcare service is not effective, is not safe, or its services are not appropriate to the standards of practice, it would be evaluated as a poor quality health service[42]. While this example may be intuitive, it shows that that these variables can serve as outcome measures for the quality of a healthcare system. However, although the definition of quality is shaped by these variables, a frame of reference must still be defined to differentiate high quality from low quality. Donabedian[40] defines quality as technical and interpersonal performances which are "*the best in practice because on average [they are] believed to produce the greatest improvement in health*." In many healthcare services, the "*best in practice*" performances are usually established in guidelines which summarize the strongest evidence on interventions to improve a patient's health outcome – a value shared by researchers, health practitioners and patients alike[40]. The dimensions of quality of PR programs are validated by the best evidence summarized in systematic reviews, guidelines and consensus statements. Therefore the variation

seen in the previous survey studies could be a reflection of the inability to follow best practice thus affecting the overall quality[25,40].

1.4.1 The Relationship between Unwarranted Variation and Quality

This section explores how unwarranted variation in certain areas of PR affects the four dimensions of quality for a healthcare service. The unwarranted variation in organization and practice may negatively affect the CIHI dimensions of: 1) quality, appropriateness, and safety; and 2) efficiency and effectiveness in PR[25,42]. These two dimensions of quality depend on the consistency of PR programs in following guidelines to improve participant health outcomes[6,42]. The CIHI dimension of quality, appropriateness, and safety in PR is important because it allows us to know whether program guidelines, such as the frequency and intensity of endurance training, are truly changing participants' health outcomes and that improvements are not due to artifacts[25].

Inconsistencies in the delivery of PR also negatively affect the effectiveness of the intervention and patient safety[25]. Although it is understood that scientific evidence is translated in clinical practice to cater to the individual needs of a patient, there remain certain procedures that should be consistently reproduced. This is because improvement in patient health outcomes is directly linked to the reproducibility of the intervention[25]. For example, despite treating patients that have different diseases, a detailed assessment prior to rehabilitation should be conducted to ensure the patient has no contraindications to exercise. The evidence supporting the clinical guidelines in PR is linked with effectively improving the health outcomes of appropriate participants despite their individual differences[1,6]. For example, the structure of PR programs should consistently

include a combination of exercise training and self-management/education. Exercise training should be performed at a certain intensity, duration, and frequency and health outcomes should be assessed using validated questionnaires. The prescribed intensity is individualized for each patient but guidelines detail how that intensity should be determined. Monitoring equipment such as pulse oximeters must be present in every setting to prevent desaturation from reaching adverse levels[1,6]. Another example regarding safety comes from a recent Canadian study, where it was reported that 20% of PR programs in Canada did not have immediate access to emergency services such as a resuscitation device or a resuscitation team[24]. These examples illustrate that if these fundamental areas are not consistently reproduced, the efficacy of the scientific evidence is threatened and participant safety is affected[25,40]

The unwarranted variation that affects the aspects of appropriateness, effectiveness and safety in turn affect participant access and participant experience in PR[6,42]. PR programs that do not have proper safety and monitoring equipment will not be able to accept eligible participants that could benefit[25]. PR programs that do not adhere to clinical guidelines and deliver ineffective interventions may negatively impact participants' experience[25,42]. In 2007, it was estimated that only 1.2% of Canadians with chronic obstructive pulmonary disease (COPD) had access to PR programs[43]. It is estimated now that only 0.4% of Canadians with COPD have access to PR programs[24]. While there are many other reasons unrelated to PR that affect access and patient experience (such as financial costs and transportation), the dimensions of appropriateness, effectiveness, and safety are important to consider [6,42].

Improving quality for PR is important because the number of PR programs and the different settings for them are expected to grow[1,6]. Most studies on PR programs do not consider evaluating program components such as team personnel, efficiency, safety, and accessibility[5]. These are practical issues that are tied to the quality of PR programs and when surveyed, were found to have a lot of variation[5,6,24]. Therefore it is important to ensure that the quality in all the programs is consistent so that all participants have a chance of achieving health benefits[1]. The first step to resolving the issues of quality control in PR programs in Canada is to identify and evaluate the current sources for variation in PR programs[5,6]. There are several general reasons for variation in a healthcare service, such as differences in organization, a lack of awareness of the healthcare service, accessibility issues, and poor adherence to guidelines[6]. However, a more detailed evaluation needs to take place in order to identify the specific sources of variation in PR programs in Canada[5,24]. Such an evaluation will allow specific components of the PR programs to be identified consistently, serve as feedback, and create international benchmarking standards[6,44].

1.5 Approaches to Measuring Quality: Quality Indicators

The integration of evidence-based practice and sophisticated technology into medicine results in it becoming increasingly specialized and complex. Practices that can be considered as common sense, such as hand-washing before and after surgery, have very low compliance rates[45,46]. Moreover, what is more concerning is that these malpractices are highly associated with increased mortality of hospital patients[45,46]. A recent study by Makary and Daniels[46] found that medical errors are the third leading cause of death in the United States. Several

organizational and individual factors were found to be associated with high rates of adverse events in hospitals[46].

Despite the amount of quality evidence and guidelines, why is there inconsistency in adhering to these guidelines? Gawande[45] claims that while current healthcare professionals are highly educated and competent, the complexity of clinical practice is leading to unintended, yet avoidable adverse events. Subsequently, Gawande[45] created the premise that "*expertise is not enough*", thus asking: "*What do you do when even super-specialists fail?*" In his seminal book, *Checklist Manifesto*, Gawande showed that using checklists significantly reduced the amount of medical errors performed by surgeons and was associated with fewer adverse events in hospitals[45,47]. His inspiration came from the aviation industry where pilots had to use checklists as the work performed by pilots was deemed "*too complex to be carried out from memory alone"*[45]. This assumption can be applied to modern medicine as well. Evidence-based medicine has generated a tremendous amount of information which clinicians are assumed to absorb and integrate into clinical practice[48]. The recent evidence provided above suggests that these challenges also apply to the field of PR.

Therefore, this *"complex high-pressure environment"* leads into two challenges that healthcare practitioners face[45]:

1) Human memory underperforms when it has to work under demanding circumstances[45].

2) There is the tendency for healthcare practitioners to rationalize mistakes because "*most of the time, they won't be critical matters"*[45].

Checklists are a solution to these challenges because: 1) checklists make the "*minimum necessary steps explicit and verifiable"*;[45] and 2) "*checklists can exist in a number of different formats and be applied to any and all situations*[45]." Thus, the purpose of checklists is to make sure that current knowledge is consistently being applied[45].

The concept of checklists can be used to identify the components responsible for the inconsistencies in healthcare services and to improve upon them. As current PR programs have issues with inconsistent practice (as described above), there is great potential in applying this concept to improve the quality of PR programs. One way to do this is to use quality indicators (QIs) [44]. QIs are statements that provide information about the quality of a specific healthcare service. These statements can point to important areas of a healthcare service that have unwarranted variation and need improvement. They are different from guidelines which are statements that facilitate healthcare professional decisions for "*specific clinical circumstances*[44]." They are also different from quality standards which are designated objectives that determine the level of care expected from healthcare professionals[44]. Guidelines and standards are the evidence upon which quality indicators are built upon and be benchmarked against[44]. Healthcare expert Donald Berwick[49] recommends that measuring the quality of a healthcare service should be in relation to the improvement it confers, and not just based on identifying its problems. This is the advantage of QIs: QIs do more than merely report the ongoing problems of a healthcare service; QIs confer the advantage of tailoring the measurements of quality in a healthcare service towards its goals for quality improvement.

1.5.1.1 Types of Quality Indicators

Under the Donabedian Framework, quality indicators in healthcare are statements for the structures, processes, and outcomes of a healthcare service[44]. There are three types of QIs used to evaluate a healthcare service. The following definitions are adapted from Campbell *et al.*[36] and Mant[44]:

Structural indicators: Evaluate the elements of the setting (equipment, staffing, and organization) in which the healthcare service is delivered.

Process indicators: Evaluate the actions and processes performed by the healthcare service for the participant.

Outcome indicators: Evaluate the impact of the healthcare service on the health outcomes of the participant.

Including both structure and process indicators allows the performance of a healthcare service to be linked to its relevant health outcomes[50]. The statements are developed upon components of a healthcare service that have a relationship with the health outcomes of the service. These statements can measure the frequency at which a specific process or outcome occurs, meaning they can be benchmarked to PR standards[44].

In summary, current research provides a strong indication of which components are integral to a quality PR program. However, recent research from the field indicates such components are inconsistently applied in clinical practice. Similar to other health services, it may be beneficial to implement QIs for PR programs to ensure that quality constituents are used on a consistent basis.

1.6 Consensus Methods for Developing Quality Indicators

The use of consensus methods can be traced back to ancient Greece[51]. The concept of these methods is to "*explore the level of consensus*" surrounding a particular topic or question[44]. In other words, consensus methods involve a process of planning wherein a group discusses and comes to an agreement on the best plan of action for a topic at hand[44]. This is very helpful in health care especially in the context of complex topics as it can provide different perspectives from each individual group member to consider before making a collective decision[51]. As Black *et al*[52]. suggest, consensus methods do not replace scientific methods of generating new knowledge. However, it is quite clear that despite the plethora of guidelines and recommendations, healthcare professionals do not make decisions solely on scientific evidence. Therefore the purpose of consensus methods is to create policy based on the best scientific evidence available and the collective experience of practicing healthcare professionals[52].

QIs are often developed using formal consensus methods. Formal consensus methods are specifically used to consider each potential QI's importance, scientific validity, reliability, and feasibility[44]. Black *et al.*[52] describe five assumptions that formal consensus methods cover:

1. Safety in numbers: The chance that a group of people would come to a wrong decision about a topic is less likely than that of a single person.

2. Authority: The group of people that are chosen to make decisions carry authority.

3. Rationality: When a group of people debate and argue over decisions, decisions improve iteratively. Reasoned argument removes decisions that have no support, and maintains decisions that have evidence.

4. Controlled processes: A facilitated and iterative process removes personal bias and peer-influence from impacting a group's decision on a topic.

5. Scientific credibility: The required scientific methods are expected in a formal consensus method.

Each of these aspects involve integrating both scientific evidence as well as an expert's experience and intuition. The integration of expert opinion reduces the limitations of scientific studies. For example, there may be limited or no studies for an important QI that is reasonable to implement, or studies may only be relevant to a small proportion of the patient population[53]. A QI derived purely from scientific evidence may not be feasible or reliable to implement. Therefore, the results from a formal consensus method can produce practical and evidence-based QIs that can be applied to evaluate PR programs. There are two general approaches for developing QIs: non-systematic and systematic approaches[44,52,54]. Non-systematic approaches are based upon case studies and accessible data. The main disadvantage of non-systematic approaches is that the development of the QIs may be based solely upon expert opinion (as the scientific literature may not be consulted) and thus, may be influenced by personal opinion. On the other hand, systematic methods try to prevent the influence of personal bias. This is achieved through synthesizing individual expert opinion with scientific evidence to establish a reliable consensus about a topic. Systematic approaches are recommended in developing indicators as they are built upon scientific evidence and may have a stronger effect in improving quality and participant health outcomes[44,55,56]. Two systematic consensus methods are the Delphi technique (Delphi)[57] and the RAND/UCLA Appropriateness Method (RAM)[58].

1.6.1 Comparing the Delphi Technique and the RAND Appropriateness Method

The Delphi and the RAM are formal consensus methods that consist of a panel rating responses anonymously for two or more rounds and receiving feedback between the rounds[57,58]. There are two features the Delphi and the RAM share. First, the anonymity of both of these methods allows panelists to answer the survey questions without being influenced by personal bias. Second, the iterative rounds in both these methods allow an opportunity for panelists to refine their answers. In comparison to the Delphi, the RAM has more advantages than disadvantages[58]. These are advantages are stated in (**Table 1.1**)

Table 1.1 Advantages and Disadvantages of the Delphi and RAM

	DELPHI	**RAM**
Size of panel	No limit to the number of panelists that participate in the survey allowing for more perspectives to be applied to a survey question, but large panel groups may be time-consuming and may yield diminishing returns and redundant information	Maximum of 15 panelists but this may limit the number of different perspectives applied to a survey question.
Use of scientific literature	Panel does not need to consult scientific literature before rating a survey question and answers may be rated solely on expert opinion.	Scientific literature is presented to the panel before rating a survey question. Panelists answer survey questions with both scientific evidence and expert opinion.
Establishment of consensus	Method to establish consensus is unclear. No cut-	Clear and systemic way to establish consensus with

	DELPHI	RAM
Establishment of consensus	off mark to remove survey questions in between rating rounds.	survey questions. Cut-off mark to remove approved survey questions in between rating rounds.
Panel discussion	Lack of face-to-face discussion may not allow panelists to listen to other panelists' perspectives.	Face-to-face discussion may allow panelists to listen to other panelists' perspectives. Panel discussion may be affected by differing personalities of panelists. For example, a particularly outspoken panelist may speak more of their own viewpoint and not let other panelists speak as much.

1.7 Rationale for Developing QIs in Canada

While QIs for PR have been developed in some countries in Europe[37,59] there currently are no QIs for PR in Canada. A reasonable question to ask is: if QIs have already been developed in some countries why are other countries, including Canada, not adopting these QIs? It seems fastidious to create separate guidelines and QIs since the literature has shown the same effect

across many studies from different countries[7]. Thus, it may seem tempting to assume that QIs are transferrable. However, that is not the case for three main reasons.

First, the two previous studies that developed QIs did not explicitly disclose the methods by which they derived the QIs[37,59]. Güell *et al.* [59] report that "*in an initial phase, each section of the document was developed by two authors of the group working independently. After the section*

was reviewed by all the authors, a second draft was prepared and the final document was put together with the consecutive revisions of the entire group, until consensus was reached." Similarly, Yohannes *et al.* reported that they developed their QI statements based on guideline recommendations from the British Thoracic Society (BTS)[37] but without further explanation on the expertise of the team or how consensus was reached.

The lack of details of the consensus process for both studies increases the possibility of bias. As QIs are tools developed from evidence-based medicine, it is important to confirm that the development of QIs follows an evidence-based protocol[42]. Despite there being a strong international framework for developing healthcare QIs, we do not know what specific criteria of evidence previous studies used to select their QIs. Güell *et al.*[59] base the creation of the QIs on expert opinion, but do not have a criterion to discriminate which QIs have more importance than the others or which ones are more feasible; Yohannes *et al.*[37] do not comment on any further methodology at all. For reasons of transparency and reliability, we believe it is important to design a study that shows the process of consensus. This will allow other jurisdictions to have a reliable and evidence-based process to decide if these QIs meet their needs[52].

QIs are not easily transferable between countries[60]. With regards to each country, there are important geographical and administrative factors that lead to the development of different QIs[60]. The evidence can be seen in a study by Marshall *et al.*[60] in 2002. They compared quality indicators developed in the United States and the United Kingdom. The important finding from this study is that many of the quality indicators differed in terms of ability for operationalization and clinical practice[60]. Only 56% of quality indicators from the United States had exact or near

equivalents to the United Kingdom. The authors concluded that QIs "*cannot be transferred from one country to another without going through a process of modification*[60]."

PR programs certainly reflect the difficulties of transferring QIs between countries to some extent[5,61]. Spruit *et al.*[5] found that the QIs developed by some countries in Europe may not be appropriate for PR programs in Canada. First, there are notable differences in the structure and delivery of PR programs in North America and Europe[5,24]. These differences are seen when it comes to the type of interventions (treadmill training versus Nordic walking training) and the type of healthcare professionals involved (the profession of respiratory therapy is not recognized in most European countries)[5]. Desveux *et al.*[24]'s study found that while the majority of countries delivery PR in outpatient clinics, 65% of PR programs in Ireland were offered in community settings. PR programs in rural and community settings are also increasing in Canada[24]. Furthermore, Canada and Ireland are two of the few countries to provide community maintenance programs[24]. As hospital clinics, outpatient clinics, and community settings differ by access and resources, certain QIs may be more important and more feasible to implement in comparison to other settings. With regards to organization, physical therapists are the primary PR staff in Sweden, the United Kingdom, and Canada while physical therapists participate in less than half of the PR programs in the United States[61]. The difference in the formal education between physical therapists and respiratory therapists (especially with regards to exercise training) would make QIs regarding staffing in the United States very different from those in other countries[1]. Lastly, a recent study by Garvey *et al.*[16] reports discrepancies in national guidelines for exercise prescription in PR. Thus, the evidence suggests that QIs developed from different countries cannot be easily transferred to one another. Each country has a different

healthcare system and different payer systems that would very likely make a single uniform consensus process long and difficult. It is impractical and unethical towards patients to delay the process of quality improvement for the purposes of uniformity when smaller jurisdictions can develop the same quality improvement protocol, albeit with some repetition, in a much shorter period.

1.8 Purpose

The purpose of this thesis is to use a modified RAM (MRAM) protocol as a consensus method to develop a list of QIs to evaluate a PR program's quality.

Chapter 2: Methodology

This study used the RAM method because this study requires both expert opinion and the available scientific evidence for answering the research question[44]. This study modified the format in which the panel rates and discusses the QIs. Panelists rated the QIs using an online survey software FluidSurveys (SurveyMonkey, Ottawa, ON), and thus for the rest of the thesis the study method will be referred as the modified RAM method (MRAM) instead of RAM. The study allowed the QIs to be created in an online survey format. This made the survey more accessible and convenient for the panel, and data collection more convenient. During the panel discussion, panelists were also given the option of adding or modifying any QIs. We believe this modification allowed us to develop a list of QIs that identifies consensus from two levels of analysis: one based on a scientific framework and one based on expert opinion[42].

2.1 Overview of the Study Protocol

A scoping review was previously conducted to develop potential quality indicators for PR programs across Canada. Prior to beginning the study, ethics approval was received from the Providence Health Care Research Ethics Board (Certificate number: H16-01678). Following the MRAM protocol, fifteen experts in the field of PR were recruited and provided informed consent. Selected panelists were given a password to access the survey. They rated the proposed QIs on a 9-point Likert scale[58] based upon four criteria (importance, scientific soundness, reliability, and feasibility). Panelists were given four weeks to rate the QIs. After the first round of rating was complete, the QIs were analyzed to determine if the panel achieved consensus. An appropriateness score for each QI was sent to each panelist for review. Panelists were able to see how their own rating score compared to that of the rest of the group. A teleconference meeting

was subsequently held to discuss the QIs that were rated as uncertain or with disagreement (see *Data Analysis 2.8* for details) in round one. The teleconference meeting was facilitated by three moderators. After the teleconference meeting, the panel received a summary of the discussions, and were given two weeks to make any changes in their ratings. After the second round of rating was complete, the rating forms were exported from the survey for analysis. At the conclusion of data analysis a final list of QIs were created.

2.2 Scoping Review

The proposed QIs were derived from a scoping review performed previously. The scoping review created an inventory of potential structural, process, and outcome indicators as well as existing quality indicators relevant to PR programs. This was done by investigating evidence-based guidelines, systematic reviews, randomized controlled trials (RCTs), and audits or technology reports. A final list of QIs was generated from this review in preparation for the RAM study.

2.3 Panel Composition

The panel was composed of a multidisciplinary group of health professionals working or conducting research in PR programs in Canada in different geographical locations, settings, and professional organizations[58]. In Canada, there are ten different healthcare disciplines typically involved in PR programs; however, only five types of healthcare professionals (respiratory therapists, dietitians, physiotherapists, nurses, kinesiologists and physicians) are involved in over 50% of the programs[24]. One to two professionals of each healthcare discipline were invited to be on the panel.

2.4 Panel Selection

The potential list of candidates for the panel was generated using several techniques. The members of the Canadian Thoracic Society (CTS) Chronic Obstructive Pulmonary Disease Clinical Assembly (research group overseeing this project) provided names of known PR experts in physical therapy, medicine, and respiratory therapy. A general mail-out to the membership of the Canadian Respiratory Health Professionals, with the inclusion/exclusion criteria, was also sent and generated several responses of interest. From these sources, panelists were chosen based on their ability to represent the specificity and credibility of opinion required to select the QIs for PR programs (**Table 2.1**)[44,58]. In addition, a past participant of pulmonary rehabilitation was invited to join the panel. After the list of potential candidates was created, candidates were sent a formal invitation letter via email explaining the rationale for this project, the details of the study, and a timetable of the study[58]. Interested candidates were asked to confirm their consent by replying by email and to send their availability as well as their curriculum vitae to provide more information about their expertise and to assist the nomination group in creating the panel[58].

Table 2.1 Inclusion/Exclusion Criteria for Expert Panel

Inclusion/Exclusion Criteria
1. Member of either Canadian Respiratory Health Professionals or Canadian Thoracic Society
2. Healthcare professionals include: physicians, respiratory therapists, nurses, physical therapists, and kinesiologists
3. Service users need to be participating in or be a recent graduate of PR program in Canada
4. Academics need to have a scientific publication relevant to PR
5. Need to be proficient in English
6. Have access to internet and email

2.5 Rating of Quality Indicators

Information about each QI was provided for the panel and included each QI's definition, quality dimension, and evidence. An online survey software, FluidSurveys (SurveyMonkey, Ottawa, ON), was used to create the QI rating form which was used to rate the QIs[62]. The criteria of each QI was adapted to match the four criteria from the healthcare QI framework developed by Kelley and Hurst: importance, scientific soundness, reliability, and feasibility (**Table 2.2**)[42,44]. On the QI form, each criterion was reworded into a question that the panelists could answer based on the 9-point Likert scale.[58]

Table 2.2 Quality Indicator Criteria

Quality Indicator Criteria	Criterion Definitions	Statement on Survey
Importance	The QI evaluates a health outcome that is impacted by the healthcare service.	Does the QI have significant impact on participant health outcomes?
Scientific Soundness	The QI is backed by strong scientific evidence and validity.	Does the QI represent what it claims to measure, and is there scientific evidence supporting the indicator?
Reliability	The QI can be measured consistently by different people, in different settings, and in different times.	Will the QI have consistent results in different settings? **Changed after panel discussion for final round:** Will the

Quality Indicator Criteria	Criterion Definitions	Statement on Survey
Reliability		measurement of this quality indicator be consistent across different settings or locations, and timeframes?
Feasibility	It is practicable to measure this QI as there are already similar and comparable measures in place.	Is it possible and practical to measure this QI for all PR programs? **Changed after panel discussion for final round:** Is this QI practical to measure in a PR program? (Note: this criterion is not asking if it is feasible to include this item in PR programs. Rather, it is asking if it is possible and feasible to measure whether or not this item takes place in PR programs).

2.6 MRAM Panel First Round

A list of instructions describing the study, the methodology, and the systematic review were first sent to the panel. Each panelist then received a password to access the survey. This initial survey had 52 pages, with two QIs per page. Each page had the four criteria definitions and definitions

for the different types of QIs. At the end of each different QI section (structure, process, and outcome) there was a page where panelists could write any comments. These comments were used to prepare the documentation for the panel discussion. The survey was also formatted such that each QI contained hyperlinks to the relevant scientific evidence (taken from the scoping review) for the given QI. For the first round, the panel were given four weeks to rate all the QIs. Reminder emails were sent out via the survey software every two weeks[62]. At the end of the four weeks the survey was closed and the panel no longer had access to change their answers. The survey exported the survey data in an Excel (Microsoft, Redmond, WA) spreadsheet format for data analysis[62]. Appropriateness, inappropriateness and the disagreement index (DI) were calculated in preparation for the panel discussion and final round of rating.

2.7 MRAM Panel Discussion and Final Round

QIs that were classified as uncertain or had a disagreement score greater than one (DI > 1) were discussed in a teleconference meeting before they were re-rated. The teleconference was conducted through Adobe Connect (Adobe, San Jose, CA) and offered an opportunity for the panel to anonymously engage in a discussion and clarify any confusion before re-rating or modifying the QIs. Three moderators facilitated the teleconference meetings. Four teleconference meetings were scheduled to enable participation of all panelists. After these meetings, a summarized transcript of all teleconferences was given to the panel members as a guide for re-rating the discussed QIs. A final survey was created in the same format as the first survey, but panelists were only required to rate the QIs they discussed. The panel had two weeks to re-rate the QIs on the final survey. Panelists rated the QIs that did not satisfy all the criteria for appropriateness and agreement in the first round.

2.8 Data Analysis

2.8.1 Calculating Appropriateness, Inappropriateness, and Disagreement

The following paragraph is adapted from Fitch *et al.*[58] and Esrailian *et al.*[63] to describe the calculations for appropriateness, uncertainty, inappropriateness, and disagreement. Two conditions must be met. The first condition is that in order for the QI to be categorized as either appropriate, uncertain, or inappropriate, the median score of the panel falls in the specific range of the Likert scale: for inappropriateness the median score falls between 1 and 3; 1-3), for uncertainty the median score falls between 4 and 6, and for appropriateness the median score falls between 7 and 9. The second condition is that the panel is in agreement with each other. The second condition is measured with a Disagreement Index (DI). A DI less than one confirms that the panel is in agreement.

Therefore, for each QI to be categorized, the following conditions are:

1) Appropriateness: A median score between 7 to 9 and a DI < 1

2) Inappropriateness: A median score between 1 to 3 and a DI < 1

3) Uncertainty: A median score between 4-6 and DI < 1

4) Disagreement: DI > 1 Disagreement between the panelists

Disagreement between the panelists suggests that the majority of the panelists are at both ends of the Likert scale. The RAM uses the DI score to resolve issues regarding bias from a small or a large panel group. The following paragraph explains how the DI is calculated (see Appendix C for formulas used in Excel (Microsoft, Redmond, WA)).

First, the interpercentile range (IPR) measures the spread of the panel scores between the 30th and 70th percentiles. A smaller IPR means a smaller spread of scores, thus more agreement and a larger IPR indicates the converse: a larger spread of scores. Nonetheless, the IPR is not enough to measure agreement because it does not take into account which side of the scale the spread of the scores is on or whether the scores are equally distributed on either end of the scale. The interpercentile range adjusted for symmetry (IPRAS) adjusts for this problem; it adjusts the IPR to show the *"broadest IPR that would constitute agreement at a certain asymmetry index*[58]*."* *Using the IPRAS therefore ensures that the measure of disagreement between the panel is "proportional to the degree of asymmetry in the distribution of responses*[63]*."* See Appendix B for the IPRAS calculation. Comparing the IPRAS and IPR scores determines whether disagreement between the panelists is occurring. If the IPR score is greater than the IPRAS score, the median and distribution of scores lies outside of the 30th and 70th percentiles even after adjusting for asymmetries. This can be seen with the DI by dividing the IPR over the IPRAS; if the DI ratio of the scores is less than or equal to one then there is agreement and if the DI ratio is greater than one then there is disagreement among the panel.

The survey was timed and immediately closed after the given date. This prevented the participants from changing their answers. The survey data was collected from the survey software and compiled into an Excel (Microsoft, Redmond, WA) spreadsheet and exported for analysis. Following the RAND/UCLA protocol[58], the median and inter-percentile range adjusted for symmetry and DI were calculated for each QI. After the calculations were made, logic functions were applied to determine QI rating. According to the logic functions, a QI criterion that satisfied a particular rating (appropriate, inappropriate, uncertain, having disagreement in the

panel) would be labeled TRUE. If it did not satisfy a particular rating, it would be labeled FALSE. A QI criterion was only considered to be properly rated if it was labeled TRUE in the row corresponding to the respective rating and at the same time labeled FALSE in the row determining if the cell had a DI > 1. Afterwards, the QIs were sorted into separate spreadsheets according to their ratings.

2.9 Results

2.9.1 Expert Panel

Fourteen Canadian PR experts were invited as panelists and 12 agreed to participate in the study. The panelists represented different healthcare professions in PR and were from different provinces in Canada. These healthcare professionals included physical therapists, physicians, academics, nurses, pharmacists and respiratory therapists). Our panel also included one PR participant – an individual with COPD who had completed a PR program approximately one year prior to the study. The years of experience our panel had working in PR ranged from 5 to 30 years. Six panelists had 5-10 years of experience, three panelists had 11-20 years of experience, and three had 21-30 years of experience in (**Table 2.3**).

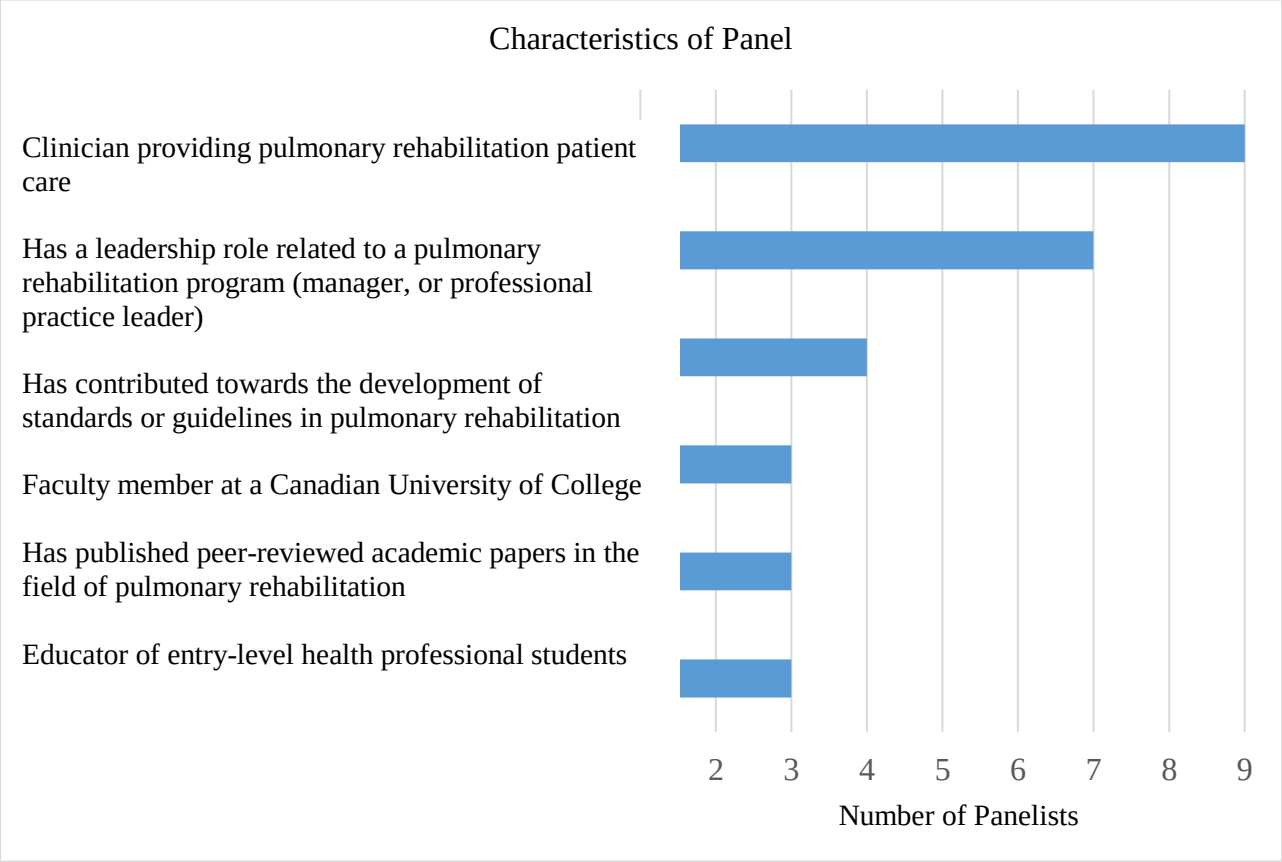
Table 2.3 Characteristics of Expert Panel

Healthcare Professions in PR*	Panelists n (%)
Physical Therapist	4 (33%)
Physician-Academic	2 (17%)
Respiratory Therapist	1 (8%)
Respiratory Therapist-Clinical Educator	1 (8%)

Healthcare Professions in PR*	Panelists n (%)
Nurse	1 (8%)
Kinesiologist-Academic	1 (8%)
Pharmacist-Clinical Educator	1 (8%)
Years of Experience in PR*	**Panelists n (%)**
5 – 10 Years of Experience	6 (50%)
11 – 20 Years of Experience	3 (25%)
21 – 30 Years of Experience	3 (25%)
*Panel also included one PR participant/patient.	

Many of the panelists had multiple roles in PR. For example, many of our panelists were clinicians, involved in creating PR-related health policy, or had published peer-reviewed research in PR (**Figure 2.1**).

Figure 2.1 Characteristics of Expert Panel



2.9.1 First Round Results

Ninety potential QIs were identified in the scoping review and presented to the panel for rating in the survey. After the first round, the panel rated 36 QIs as appropriate and 54 QIs as uncertain or with a DI > 1. Of the different types of QIs, 45% of the proposed structural QIs, 36% of the proposed process QIs, and 47% of the outcome QIs were rated as appropriate (**Table 2.4**). Of the 54 QIs rated as uncertain or with a DI > 1, the majority of QIs were rated uncertain in reliability, followed by feasibility, scientific soundness, and importance (

Table 2.5).

Table 2.4 First Round Rating of QIs (Type)

Type of QI	Total Proposed QIs, n	First Round QIs Rated Appropriate, n (%)	First Round QIs Rated Uncertain and/or a DI > 1, n (%)
Structure	22	10 (45%)	12 (55%)
Process	53	19 (36%)	34 (64%)
Outcome	15	7 (47%)	8 (53%)
Total	90	36 (40%)	54 (60%)

Table 2.5 First Round QIs (Criteria) Uncertain and/or DI > 1

QI Criteria	First Round QIs Rated Uncertain and/or a DI > 1, n (%)
Importance	13 (24%)
Scientific Soundness	16 (30%)
Reliability	50 (93%)
Feasibility	41* (76%)
Total QIs	54 QIs
Note: Multiple criteria in a single QI could be rated as Uncertain, Inappropriate, or with a DI > 1 *Feasibility also had 2 QIs with a DI > 1	

2.9.2 Panel Discussion

Four different teleconferences over the span of two weeks were scheduled for the panel to discuss the 54 QIs rated as uncertain or with a DI > 1. Ten out of the twelve panelists participated in the discussions. The purpose of the discussions was not to reach consensus, but

rather to provide each panelist with other perspectives they may not have considered when they were rating in the first round. Each discussion was two and a half hours long and held between two to four panelists. This discussion format allowed each panelist to articulate their thoughts without any pressures regarding time or from their peers. The panelists also had the opportunity to clarify any misconceptions regarding the criteria definitions and the QIs. The teleconferences were recorded and summarized, and a document with compiled summaries of the four teleconferences was provided to the panelists. This was to aid them in the final round of rating.

There were several issues brought up by the panelists during the discussion sessions. These are summarized in **Table 2.6**. To summarize, the panel discussed the following: criteria used to rate the QIs, the QIs chosen for inspiratory muscle training, outcome QIs, the rewording of QIs, QIs relating to PR staffing, and QIs relating to participant follow up.

Table 2.6 Summary of Panel Discussion

Summary of Panel Discussion	
Criterion Definitions	Many of the panelists noted that they were confused about the definitions of feasibility and reliability during the first round of rating. 1. **Feasibility:** Many panelists thought that feasibility meant whether it was feasible to provide the care specified by the QI in a PR program. During the discussions, it was clarified that feasibility meant whether it was feasible to measure if the

Summary of Panel Discussion	
Criterion Definitions	particular QI is present in a PR program. 2. **Reliability:** Many panelists thought that reliability meant whether it was reliable to provide the care specified by the QI in a PR program. During the discussion, it was clarified that reliability meant whether the particular QI could be measured consistently across different PR settings during different time periods. After these definitions were clarified, the panel agreed that rating the reliability and feasibility of the QIs became clearer.
Inspiratory Muscle Training	The entire panel agreed that the QIs regarding inspiratory muscle training did not satisfy the QI criteria of importance, scientific soundness, reliability, and feasibility. The QIs regarding inspiratory muscle training are as follows: S13 Respiratory Muscle Assessment and Monitoring Equipment The pulmonary rehabilitation program uses validated equipment for respiratory muscle assessment. S14 Respiratory Muscle Training Equipment

Summary of Panel Discussion	
Inspiratory Muscle Training	The pulmonary rehabilitation program uses validated equipment for inspiratory muscle training. P37 Respiratory Muscle Assessments Baseline respiratory muscle assessment is done according to a standardized test. P38 Inspiratory Muscle Training Intensity Participants begin their inspiratory respiratory muscle training at a PiMax of at least 30% and increase load as tolerated. P39 Inspiratory Muscle Training Frequency PR participants train their respiratory muscles at least 5 times a week, including home training. P40 Inspiratory Muscle Training Duration Participants train their respiratory muscles at least 10-15 minutes daily. P41 Inspiratory Muscle Training Exercise Progression The respiratory muscle training program is progressed as tolerated.

Summary of Panel Discussion	
Inspiratory Muscle Training	O15-Respiratory Muscle Strength Outcome Measures Respiratory muscle strength outcomes are assessed after the PR program.
Outcome Measures	Many panelists suggested that several outcome QIs would have low reliability because program audits could occur before programs had ended and collected outcome measures. Other outcome QIs and also had issues with scientific soundness. The QIs and comments regarding outcome measures are as follows: O3-Fatigue Outcome Measures Fatigue outcomes are assessed after the PR program. Comments by panelists: "*In our area, every PR program uses the chronic respiratory questionnaire that has been validated and one of the domains is fatigue."*

Summary of Panel Discussion	
Outcome Measures	"*Fatigue isn't something I work on as usual. I use the CRQ which also includes fatigue in the measure, as time goes on perhaps we will use that measurement tool as well."* O7-Physical Activity Outcome Measures Physical activity outcomes are assessed after the PR program. Comment by panelist: "*This is really big work, and how to really measure outcomes in a consistent fashion? If we want reliable measurement, we may have to indicate how that outcome will be measured, how will the auditor evaluate this program at different levels and some indication that measurement took place.*" O8-Psychological Outcome Measures Psychological outcomes are assessed after the PR program. Comment by panelist: *"I would suggest removing the "s" at the end, if there is a (one)*

Summary of Panel Discussion	
Outcome Measures	*fatigue outcome assessed. I think the confusion could be from answering yes or no to this. How comprehensive does it have to be to answer yes?"* O9-Health Care Utilization Outcome Measures Healthcare utilization outcomes are assessed after the PR program. Comments by panelists: *"Health care utilization means a lot of things. What would be measured? It could be community RTs, doctor visits, ER visits, increase use of medication, more costs to healthcare systems*." "A support of visits to GP or the issue may be a medical non-pulmonary issue. Is it amendable to the activities and education provided in PR program?" O11- Composite Outcome Measures Composite outcomes are assessed after the PR program. *"Composite sounds vague. PR doesn't have a big impact on*

Summary of Panel Discussion	
Outcome Measures	*mortality so might not be important to measure it (scientific soundness). Some PR programs do track mortality, and was measured because the medical director suggested that they track mortality to compare to patients who don't participate in PR. However this is very difficult to do when looking for matched controls. Some people may think PR may not have a big impact on mortality, and therefore may be why some of us don't think it is important to measure."* *"Have to plead ignorance and say I wasn't sure what the composite measure was. When staff don't know about it, it is reasonable why importance may be rated low. These are often more prognostic for mortality, and in rehab we perhaps are not looking at prognostic and looking at more positive things."* O13-Muscle Strength Outcome Measures Muscle strength outcomes are assessed after the PR program. Comments by panelists:

Summary of Panel Discussion	
Outcome Measures	*"People are using different tools to measure and this makes the feasibility and reliability of this QI challenging."* O14-Muscle Endurance Outcomes Measures Muscle endurance outcomes are assessed after the PR program. *"People are using different tools to measure, does this make the feasibility /reliability of this QI challenging. Muscle endurance is not being specifically measured, and there may be significant discrepancies in the measures of muscle strength which makes it quite difficult to measure across different clinics."*
Rewording QIs	Many panelists suggested 23 QIs should be reworded to improve their clarity (see Appendix D 1) Several panelists stated that the word "*should*" affected the reliability of the QI. The criticism the panel had was that this QI implied that it was optional for exercise testing and training to be supervised and was therefore an unreliable QI.

Summary of Panel Discussion	
Rewording QIs	The reworded QIs and related comments are as follows: S20 Team of Essential Staff: Exercise testing and training should be supervised by a physical therapist or a physician that has expertise in testing and training. Comments by panelist: *"Goes back to wording, "should be supervised" so what if it is not supervised. If you said it must be supervised , I can see how it will be reliable. If you don't have a physical therapist, and a kinesiologist, you may always fail on this QI. The wording of "should be" threw me off about it whether it will be a reliable QI."* 2) Another example is making some outcome QIs singular instead of plural as it affected the reliability and feasibility of the QI. Examples are presented below. Comments by panelist:

Summary of Panel Discussion	
Rewording QIs	*"I would suggest removing the "s" at the end, if there is a (one) fatigue outcome assessed. I think the confusion could be from answering yes or no to this. How comprehensive does it have to be to answer yes?"* 3) Another example is QI P16, where several panelists stated the word "comprehensive" was affecting the reliability of the QI. P16 Education: In addition to exercise training, the pulmonary rehabilitation program has a comprehensive education/self-management component. Comments by panelists: *"Perhaps there is an issue with the definition of comprehensive. What may be considered comprehensive to one group is not comprehensive to another."*

Summary of Panel Discussion	
Rewording QIs	*"If you removed the word comprehensive, it will make the QI more reliable and more clear."*
PR Program Staffing	The panel stated that the following two QIs were difficult to measure reliably as many PR programs had different multidisciplinary teams. The QIs and comments regarding PR program staffing are as follows: S19 (Team of Essential Staff: Program has an essential team of a physical therapist and respiratory therapist with access to a physician and a nurse. The staff may be full time or provide education or consultation in the program). S20 (Team of Essential Staff: Exercise testing and training should be supervised by a physical therapist or a physician who have expertise in testing and training).

Summary of Panel Discussion	
PR Program Staffing	Comments from panelists: *"It depends on the validated tools to measure, unless you know the PT or physician are trained to use that validated tool to record testing and or training and the results, you can't determine that measurement is going to be reliable. It also depends on how it is recorded…ex 6MWD, physio assistant do a test if pre-screened, would they co-sign they'd have to look at protocols and just sign it? This affects the reliability."* *"Goes back to wording, "should be supervised." So what if it is not supervised? If you said must be supervised it, I can see how it will be reliable. If you don't have a physical therapist, and a kinesiologist, you may always fail on this QI. The wording of "should be" threw me off about it whether it will be a reliable QI."*
QIs Relating to Participant Follow Up	Many panelists stated that the following three QIs would not be feasible to measure as maintenance and healthcare utilization occurred outside the PR program and obtaining patient data would be very difficult for auditors.

Summary of Panel Discussion	
QIs Relating to Participant Follow Up	The QIs and comments regarding participant follow up are as follows: P45 (Maintenance and Follow-Up: The PR program has a maintenance program that provides continued exercise training and education for the participant). Comments from panelists: *"Different types of maintenance, are they equivalent? This program can be very different from one program to another. Where is this maintenance program taking place? The issue with the maintenance programs is that they no longer reside within the PR programs. Most of them are off site and will make it difficult to track participants."* P46 (Maintenance and Follow-Up: Program follows up with patient after discharge via telephone to discuss participant's exercise program and answer questions). Comments from panelists:

Summary of Panel Discussion	
QIs Relating to Participant Follow Up	*"Normally very difficult to do unless you are in a well-resourced PR program. It is something we are supposed to do, but it is very difficult to measure. If a clinician is doing a telephone follow-up, then during each phone call, they may vary what is being said. It is difficult to determine that programs that do adhere to this do better in their outcomes in comparison to programs that don't."* *"If you go back to the previous question, P45. If a patient is being referred to a maintenance program, why the need to follow up? We are not measuring if the patient acted on follow-up, but that PR programs do that in the case that they catch patients falling through the cracks. The quality indicator is saying that are we offering the follow-up and not as much as seeing how the patient responds."* P53 (Healthcare Utilization: The pulmonary rehabilitation program assesses healthcare utilization before and after pulmonary rehabilitation, or if hospitalization occurs during the pulmonary rehabilitation program). Comments from panelists:

Summary of Panel Discussion	
QIs Relating to Participant Follow Up	*"Concerned about feasibility, in terms, it would be a problem of reliability and feasibility. Is it reliable and feasible? Even exacerbations have variation in how they are recorded. How consistently people will be doing this?"* *"Not sure what level of clearance would be able to access this info. May not be feasible for auditing."*

In this study, we also have two comments made by the patient-panelist. These two comments, we believe, were congruent with the remarks made by the healthcare professional panelists.

The QIs and comments for this example are as follows:

1) P14 Goal Setting: In consultation with pulmonary rehabilitation staff, participants develop written goals for the program

Healthcare Professional Panelist: *"[This] brings up a scenario of why the indicator might be difficult to measure. This may not end up in the medical record. Participants may not know enough about the program to know what goal that they want to reach by the end of the program. Certain patient populations may not work for each individual group because some of the patients may not respond to them. Some patients have mild cognitive impairment and may not understand the SMART goal-setting principles. That is, this QI may not be feasible from program to program and patient to patient."*

Patient-Panelist: *"I don't recall in the PR program of writing down goals, not sure whether to be able to write down goals without professional help and what goals to contemplate. It became obvious to me that I want to be active and I don't want to end up on supplemental oxygen. But I never wrote these down."*

2) P52 Home Exercise Program: The pulmonary rehabilitation program provides exercise prescription or physical activity guidance for participants to exercise at home

Healthcare Professional Panelist: *"[I am] wondering if just general guidelines are being discussed, or if it is written out and reporting verbally what they did on the program days or on the home-exercise piece. [This QI] appears hard to be reliably looked at. The only thing I would be able to show to an auditor is a slide on what participants would do after they graduate [from the program]. That is the extent of the information I am able to give them."*

Patient-Panelist: *"When I was going from different programs, I don't remember ever having a different exercise prescription. As far as my own physical activity at home, I learned on the go and nothing was written down. I remember the breathing exercises and flexibility."*

2.9.3 Final Round Results

The full process of developing the QIs can be seen in **Figure 2.2**. Before the final round of rating one panelist dropped out of the study leaving eleven of the original twelve panelists to re-rate the 54 QIs. After the panel discussion and final round of rating, the panel rated 56 of the 90 QI as appropriate. Thirty-four of the 90 QIs had a rating as inappropriate, uncertain and/or with a

DI > 1 in one of the four QI criteria. These 34 QIs were consequently discarded. These results can be seen in **Table 2.7**. Of the QIs rated appropriate in the final round, 34% are structure QIs, 52% are process QIs, and 14% are outcome QIs.

Figure 2.2 Development of QIs

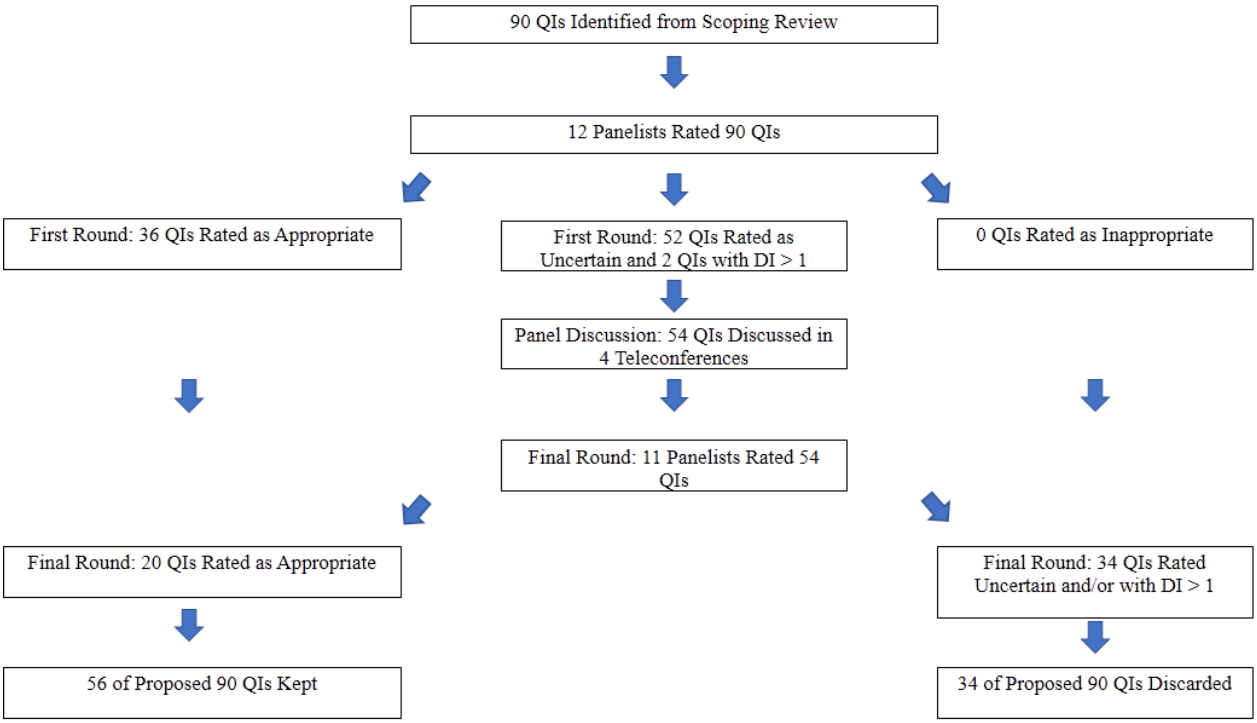


Table 2.7 Comparing Appropriate QIs (Type) Between First and Final Rating Rounds

Type of QI	First Round Rated Uncertain and/or DI > 1, n (%)	First Round QIs Rated Appropriate, n (%)	Uncertain and/or DI > 1 QIs Rated Appropriate in Final Round, n (%)	Final Round Rated Appropriate, n (%)	Final Round Discarded (Rated Inappropriate, and/or Uncertain and/or DI >1, n (%)
Structure	12 (22%)	10 (28%)	9 (45%)	19 (34%)	3 (9%)
Process	34 (63%)	19 (53%)	10 (50%)	29 (52%)	24 (71%)
Outcome	8 (15%)	7 (19%)	1 (5%)	8 (14%)	7 (20%)
Total QIs	54	36	20	56	34

When comparing the criteria ratings of the QI discussed by the panel between the first and second rounds of rating, the reliability and feasibility scores improved (became rated as appropriate in the final round) to 54% and 32%, respectively. In contrast, the QI criteria of importance and scientific soundness scores improved by 15% and 13%, respectively (**Table 2.8**). This could be explained by the fact that the definitions of the QI criteria were clarified. Several panelists understood reliability as whether it was possible for a PR program to reliably deliver a process, or measure an outcome, rather than whether the QI could be measured similarly by different auditors and in different PR settings. Panelists also understood feasibility as whether it was possible for a PR program to deliver a process, or measure an outcome, rather than whether it was feasible for the QI to be evaluated. These clarifications may explain why the ratings of uncertainty, and a DI > 1 improved between the first and second rounds of rating.

Table 2.8 Comparing QI Criteria Rated as Uncertain and/or DI > 1 between First and Final Rating

QI Criteria	First Round Uncertain and/or DI > 1 n (%)	Final Round Inappropriate and/or Uncertain and/or DI > 1 n (%)
Importance	13 (24%)	11 (32%)**
Scientific Soundness	16 (30%)	14 (41%)
Reliability	50 (93%)	23 (68%)***
Feasibility	41 (76%)*	28 (82%)****
Total QIs	54 QIs	34 QIs

Note: Multiple criteria in a single QI could be rated as Uncertain, Inappropriate, or with DI > 1.
Note: 1 QI criteria could be rated either as Uncertain and with DI > 1, or Inappropriate with DI > 1.
* 2 QIs with a DI > 1
**1 QI rated Inappropriate
***4 QIs with DI > 1, 1 QI rated Inappropriate
****5 QIs with DI > 1, 1 QI rated Inappropriate

Of the 54 QIs discussed by the panel, 23 were subsequently reworded. 43% of the reworded QIs were then rated in the final round as appropriate, with the remaining QIs discarded. Of the reworded QIs the criteria of feasibility and reliability were rated as appropriate in the final round to a greater degree in comparison to the criteria of importance and scientific soundness. Other than the aforementioned reason of clarifying the criteria definitions, the panel found the words, *"should"*, and *"comprehensive"* problematic to the reliability and feasibility of the QIs. The panel also argued that including the word "*and*" in the QIs related to PR staffing, and omitting the phrase, "*pre and post PR*" in certain outcome QIs affected the QIs' reliability and feasibility. These are some possible reasons for the trend seen in **Table 2.10**. The specific number and type of reworded QIs re-rated as appropriate and discarded can be seen in **Table 2.9** and **Table 2.10**.

Table 2.9 Comparison of the Type of Reworded QIs Between Discussion and Final Rating

Type of QI	First Round Rated Uncertain and/or DI > 1 n (%)	Reworded QI After Panel Discussion n (%)	Reworded QI Rated Appropriate in Final Round n (%)	Reworded QI Rated Discarded in Final Round (Rated Inappropriate, and/or Uncertain and/or DI >1 n (%)
Structure	12 (22%)	5 (22%)	4 (40%)	1 (8%)
Process	34 (63%)	10 (43%)	5 (50%)	5 (38%)
Outcome	8 (15%)	8 (35%)	1 (10%)	7* (54%)
Total QIs	54	23	10	13
*3 QIs with a DI > 1				

Table 2.10 Comparison of Reworded QIs' Criteria Between Discussion and Final Rating

QI Criteria	QIs Reworded After Panel Discussion n (%)	QIs Reworded Rated Appropriate in Final Round n (%)	QIs Reworded Discarded in Final Round (Rated Inappropriate, and/or Uncertain and/or DI >1) n (%)
Importance	4 (17%)	1 (25%)	3 (75%)
Scientific Soundness	6 (26%)	2 (33%)	4 (67%)
Reliability	19 (83%)	12 (63%)	7* (37%)
Feasibility	20 (87%)	9 (45%)	11** (55%)
Total QIs	23	10 (43%)	13 (57%)
Note: Multiple criteria in a single QI could be rated as Uncertain, Inappropriate, or with DI > 1. Note: 1 QI criteria could be rated either as Uncertain and with DI > 1, or Inappropriate with DI > 1. *(1 QI with DI >1) **(3 QIs with DI >1)			

This study developed a set of QIs that could cover the various educational and exercise components of a PR program. The number of QIs relating to each PR component can be seen in **Table 2.11**. Of the nine PR components, the structure QIs covered five components, whereas the process and outcome QIs covered three components and one component respectively.

Table 2.11 PR Components for Rated Appropriate

PR Components for Structure QIs Rated Appropriate	**Number of QIs n(%)**
Exercise Training and Assessment	7 (13%)
Space and Access	4 (7%)
Staffing	3 (5%)
Safety Equipment	3 (5%)
Other	2 (4%)
PR Components for Process QIs Rated Appropriate	
PR Outcomes Assessment	13 (23%)
Exercise Prescription	9 (16%)
PR Program Organization	7 (13%)
PR Components for Outcome QIs Rated Appropriate	
Outcome Measures	8 (15%)

2.10 Discussion

This study developed 56 QIs using a MRAM method, a method that synthesizes scientific literature and expert opinion. The final 56 QIs comprehensively cover the organization and delivery of a quality PR program. The QIs were organized according to the Donabedian Framework, resulting in 19 structural QIs, 29 process QIs and 8 outcome QIs. The structural QIs cover areas of safety equipment, exercise training equipment, exercise testing equipment and staffing. The process QIs cover areas of exercise prescription, exercise testing, and outcome evaluation. Finally, the outcome QIs cover evidence-based outcome measures used in PR.

2.10.1 Strengths

There are several strengths to this study. First, the MRAM satisfies the five assumptions made by Black *et al.*[52] (**Table 2.12**)

Table 2.12 Comparison between Black et al.'s [52] Assumptions and MRAM Methodology

Assumptions	MRAM Methodology in this Study
Safety in numbers: The probability that a group of people come to a wrong decision about a topic is less likely than that of a single person.	This study had a panel of twelve experts in PR that hold various leadership positions in academia and clinical practice. This number is similar to other RAM studies developing QIs [53,64–66].
Authority: The group of people that are chosen to make decisions carry authority.	Several of the experts have also contributed guidelines and policy in the field of PR, thus there is an element of authority supporting the QIs selected.

Assumptions	MRAM Methodology in this Study
Rationality: When a group of people debate and argue over decisions, decisions improve iteratively. Reasoned argument removes decisions that have no support, and maintains decisions that have evidence.	The panel discussion allowed for argument, based upon clinical experience and the scientific literature, to occur in a controlled process.
Controlled processes: A facilitated and iterative process removes personal bias and peer-influence from impacting a group's decision on a topic.	The discussions were facilitated by three moderators who ensured that all the panelists had the opportunity to provide their thoughts on a particular QI.
Scientific credibility: The required scientific methods are expected in a formal consensus method.	Lastly, the format of the survey in which panelists had access to the scientific literature for each QI and were clarified on the definitions of the QI criteria allowed this study to operate with scientific credibility.

In this study, the criteria that had the fewest ratings of uncertainty and/or a DI > 1 were the criteria of *importance* and *scientific soundness*. The criteria that had the most ratings of uncertainty or a DI > 1 were the *feasibility* and *reliability* criteria. This suggests that the panel agreed that the scientific evidence and importance of the QIs was established, yet experienced difficulty in choosing which QIs are reliable and feasible to measure.

2.10.2 Comparisons to Other QI Studies' RAM Methodology

This study's methodology was similar to other studies that used the RAM. Previous QI studies had similar panel sizes, used two rounds for rating, and used the DI to calculate disagreement between the panel[53,65–67]. Previous studies[42,53,63,66,67] also used a framework to rate their QIs, although the criteria they used differed slightly from the criteria recommended by Kelley and Hurst[42], and often did not cover the four rating criteria (importance, scientific soundness, reliability, and feasibility). Although the definitions are similar with regards to content, there are differences in the definition names. For example, the study by To *et al.*[53], developing QIs for asthma treatment used criteria of "*relevance*", "*room for improvement*", "*overall*", "*validity*", *and* "*feasibility*", while Shekelle *et al.*[64], developing QIs to evaluating the care of vulnerable elderly individuals, used "*validity*" (Shekelle *et al.*[64] subset validity into three separate definitions). Both To *et al.*[53] and Shekelle *et al.*[64] have criteria (both termed "*validity*") that are similar to this study's definition of scientific soundness: there is scientific evidence supporting the QI. To *et al.'s*[53] criteria of relevance (the QI can improve the quality of care provided by a healthcare service), room for improvement (the QI is sensitive to problems in a healthcare service), and overall (the QI can evaluate the quality of care provided by a healthcare service) is similar to our study's criteria of importance. Contrary, Shekelle *et al.'s*[64] criteria of validity differs from this study's criteria of importance and To *et al.'s*[53] similar criteria as one of their definitions of validity is that adhering to the QI suggests that a high quality of care provided by the healthcare service. Additionally, there are some differences between this study's criteria and To *et al.'s*[53] criteria. With regards to feasibility, To *et al.*[53] state that data around the QI can be collected in a precise manner, adjusted appropriately (with regards to different disease severity and risk factors), and collected for different patient populations (adult, child, and sex). This

definition of feasibility seems to fit more precisely with our study's criteria of reliability (To *et al.*[53] did not use reliability as a rating criteria, as did Shekelle *et al.*[64]). To *et al.*[53] seem to be asking their panel whether the QI can be consistently assessed by auditors over different patient populations, and settings. While the criteria definitions between our study is similar between To *et al.*[53] and Shekelle *et al.*[64], there are differences (as provided by the examples above) that may lead to some discrepancy in interpretation. It may be more beneficial for future QI studies using the RAM to follow the framework provided by Kelley and Hurst[42] for consistency and clarity.

Additionally, there are two notable differences with this study's methodology and the other QI studies. The first difference is the inclusion of a patient-panelist in the expert panel. Including a patient who uses the healthcare service offers a unique and important perspective that healthcare professionals may not have. There was a comment made by a patient-panelist that we believe was unique to the patient experience.

Referring to QI P11-Psychological Assessment Test, the patient-panelist stated "*[it is] really tough because it is so individualized, I don't remember any time that I was evaluated. But the staff there was so expert, that I don't think they don't need to formalize a psychological assessment. The trick is not to get staff to re-do tests over and over again for the sake of auditing, because this comprises the logistics of a program. Sometimes the problem with the auditor, is that you can audit yourself to death."*

This comment provided a perspective that was not brought up by the other panelists, who were healthcare providers. The patient noted that at times, researchers and clinicians may be pre-

occupied with performing certain measurements in a PR program. This may result in compromising the participant's program preferences, leading to an unsatisfying experience with the healthcare service. Including patients in the development of QIs, therefore, may help identify QIs that satisfy or may not satisfy the CIHI quality dimension of patient experience (as we have provided in the results above). Interestingly, the RAM protocol does not require a patient-panelist, a weakness noted by van Engen-Verheul *et al.*[67]. These findings support the recommendation that future QI studies using the RAM should include a patient-panelist on their expert panel.

The other main difference between this study's RAM methodology and the other studies is the recording and transcription of the panel discussion. Other QI studies did not report what their panelists said during their discussions[53,65–67]. Use of the recordings was particularly helpful in our study because it provided explicit feedback to the panelists regarding the QIs that were rated as uncertain or with a DI >1. The list of QIs discussed by the panel was very long and it was often difficult to remember what different panelists said about a particular QI. Therefore, having a method to record and provide a summary of main points for the QIs discussed was very valuable. For this study, the qualitative data obtained from the QI discussion provided a possible explanation as to why a large amount of QIs with initial ratings of uncertainty and a DI >1 for feasibility and reliability were later rated as appropriate in the final round.

2.10.2.1 Comparison of Methodology to Other PR QI Studies

The necessity for quality criteria to evaluate the delivery of PR programs around the world is recommended by the ATS/ERS[6]. Two other studies, Guell *et al.*[59] and Yohannes *et al.*[37] created

QIs for PR programs in their respective national healthcare jurisdictions. Both these organizations developed their QIs based on expert opinion and/or existing guidelines. In comparison with QIs from SEPAR and BTS, the QIs in this study were developed from a systematic and established methodology[44]. This study used a MRAM and applied a healthcare framework to develop 56 QIs. Use of the MRAM provides a methodological strength to this study's QIs as opposed to the way QIs were developed by Yohannes *et al.*[37] and Güell *et al.*[59].

2.10.2.2 Comparison of Results to Other PR QI Studies

In the QI studies by Yohannes *et al.*[37] (from the BTS organization) and Güell *et al.*[59] (from the SEPAR organization) there may be some ambiguity associated with the QI statements. For example, Yohannes *et al.*'s[37] second QI states that *"the pulmonary rehabilitation program is delivered by a multidisciplinary team*[37]*."* The QI's additional guidance note says to "*denote 'only partially met' if your rehabilitation programme has contributions from only the physiotherapists and respiratory nurses*[37]". It does not define what an actual multidisciplinary team in a quality PR program would consist of. Would having three staff consisting of a nurse, physical therapist, and respiratory therapist indicate a multidisciplinary program? Or does a team consisting of a nurse, a pharmacist, and a psychologist, whose disciplines don't include exercise supervision, qualify as a multidisciplinary program? In addition, there may be quality PR programs run by a physiotherapist and respiratory nurse that may only qualify as partially meeting this QI. The lack of clarity in this QI makes it difficult to reliably and feasibly measure. Conversely, this study developed three separate QIs related to staffing that may be more reliable and feasible to measure. Reliability and feasibility may be established because the staff which have essential roles versus auxiliary roles are clarified. In addition, satisfying these three QIs also satisfies the

definition of a multidisciplinary team that Yohannes *et al.* [37] define. Another example of ambiguity from the Yohannes study relates to assessment processes. Their eleventh QI states that *"measurements such as spirometry, exercise and health status are recorded before and after pulmonary rehabilitation."* Using the wording *"such as"* is problematic as it means that any one of these possible measures could satisfy the QI. This suggests that even though a quality program satisfies the QI, it may not evaluate important outcomes such as HRQoL and exercise capacity. This study's QIs explicitly separated each of these outcome measures in order to emphasize what a quality PR program should measure. The wording in Güell *et al.*'s [59] study also contains issues of ambiguity. The study by Güell *et al.* [59] included the word *"should"*, for many of their QIs (24/45). The problem that occurs with the use of *"should"* is that it makes the QIs indistinguishable from guidelines and standards. As previously stated, guidelines are recommendations to aid clinical practice and decisions, whereas QIs are components of clinical practice that provide a measurement of quality. Campbell *et al.*[44] uses the following example to illustrate the difference between a guideline and a QI:

Guideline: *"If a blood pressure reading is raised on one occasion, the patient* ***should*** *be followed up on two further occasions within x time."*

Quality Indicator: *"Patients with a blood pressure of more than 160/90 mm Hg have had their blood pressure re-measured within 3 months."*

This example makes it quite evident that indicators are different than guidelines and therefore should be written as such. The inability to clearly demarcate an indicator from a guideline can lead to issues of ambiguity when used in an audit. The initial 90 QIs presented to the panel in first round also used the word *"should"*; however the world *"should"* was subsequently removed after strong suggestions by the panel in the panel discussion. The panel's main

objection to the word *"should"* was that it may affect the reliability of the QI. For example, one of the QIs, S20-Team of Essential Staff, previously stated *"exercise testing and training should be supervised by a physical therapist or a physician who have expertise in testing and training."* This study subsequently reworded this QI to state *"exercise testing and training are supervised by a physical therapist or a physician who have expertise in testing and training."* Many of the QIs in Güell *et al.*'s [59] study use the word *"should"*, and may have a problem of ambiguity if they are used during an audit. Overall, the lack of clarity from Yohannes *et al.*[37]'s and Güell *et al.*'s [59] studies make their QIs difficult for auditors to interpret when evaluating a PR program. We believe the QIs from this study do not have the same issues and will be more reliable and feasible for auditors to measure in the future.

2.10.3 Comparisons to Other Approaches to Developing Quality Criteria

The ATS/ERS[1] consensus statement suggests developing both performance and process metrics for quality control of PR programs.

ATS/ERS[6] Recommendations:

1. "*PR programs should follow relevant evidence-based clinical guidelines and demonstrate the measurement of standard outcomes to document benefits, quality, and safety."*

2. "*A core set of processes and outcomes should be established to enable national and international benchmarking in PR; this should include both process and performance metrics to enable recommendations for international standards based on evidence and best practice."*

Our study did not develop performance metrics for the following reason. Performance metrics, as some PR programs in Canada currently use, may be unable to identify problems as they only measure the ongoing processes of a health service without judging whether the processes are aligned with the dimensions of quality[44]. Performance metrics used in Canadian PR programs, such as completion rate, dropout rate, and participant satisfaction are "*statistical devices for monitoring care provided to a population*" and do not provide any information about the dimensions of safety, appropriateness, or effectiveness of the program[44]. Thus, these performance metrics cannot make any specific judgments of quality and are unable to fulfill the recommendations put forward by the ATS/ERS consensus statement.

2.10.4 Applicability of Results

One of the key functions of QIs is to evaluate the ongoing knowledge translation (KT) between research and clinical practice[26]. In KT, QIs function as knowledge tools and help inform PR stakeholders and healthcare professionals of their decision-making and outcomes[68]. Within a healthcare service, Graham *et al.*[68] identified the key actions for patients to derive benefit from the healthcare system: the scientific literature must be tailored in a manner to help patients improve their health outcomes; and an evaluation system must be in place to provide feedback to improve the healthcare service. It is well recognized that there is strong evidence for the efficacy of PR and that PR programs are structured to improve patient health outcomes[5,6]. However, there are also concerns with the quality of care provided by healthcare professionals working in PR and the delivery of the healthcare service[5,6]. Therefore, PR requires the establishment of an evaluation system. The QIs developed from this study can have several applications. First, they can be used for program evaluation, benchmarking, and accreditation. Second, they can be used

towards continuing education for PR healthcare professionals. Finally, they can be used for increasing the health literacy and adherence for eligible PR participants.

2.10.4.1 PR Program Evaluation, Benchmarking and Accreditation

The development of QIs can be operationalized for auditing of programs and setting the benchmarking standards the ATS and ERS are striving for[5,6]. In addition, QIs can be used as the foundation for PR program accreditation. They may increase physician and patient awareness of the processes and benefits of PR, and thus increase physician referral rates to PR[26].

Benchmarking standards for within-program and between-program comparisons can also be established. Within-program comparisons allow a PR program to evaluate and improve itself over time. They may become a basis for continuing education of PR healthcare professionals, as discussed below. Between-program comparisons may allow for evaluations to be made within a geographical region or a particular PR setting. However, comparing QIs between PR programs may be difficult as resources and participant characteristics differ; comparisons may therefore not reflect a difference in quality in delivery[5,53]. Nonetheless, despite differing resources and participant case-loads there are certain elements within a PR program that should not differ per setting such as safety equipment, essential staffing, and exercise training principles. The QIs developed in this study cover these fundamental characteristics of a PR program and the benchmarks established therefore reflect a quality that programs should aim for.

Over time, these QIs may be used to assist healthcare monitoring systems to understand the effect new healthcare policies and research have on the quality of PR programs[53]. This is because

these QIs provide a consistent terminology for program evaluation and comparison. A study by Camp *et al*.[24] demonstrated that Canadian PR programs do have performance evaluation measures, but that these measures may drastically differ between programs due to differing definitions of their measures. The healthcare monitoring systems can be used for program accreditation. Accreditation qualifications could involve satisfying the benchmarks of our QIs. Having an accreditation process may standardize the quality of programs and improve the recognition of PR programs for eligible patients, referring healthcare professionals, and stakeholders.

2.10.4.2 Continuing Education for PR Healthcare Professionals

Ecceles *et al.*[69] show that that when healthcare professionals self-evaluate their adherence to clinical guidelines, they overestimate how well they perform by 20-30%. Moreover, contrary to conventional belief, another systematic review found that physicians who are older and have more clinical experience actually have less "*factual knowledge*," are less likely to follow medical standards[70]. This could lead to the healthcare professional practicing in a manner that is contrary to the scientific evidence. Moreover, this form of care may be associated with worse patient outcomes[70]. Additionally, educational and clinical training backgrounds may differ between the healthcare professionals involved in a particular PR program. For example, a respiratory therapist will not have as much clinical training in exercise testing in PR compared to a physical therapist, but may have more experience in educating respiratory patients. This healthcare professional may be prepared for working in collaboration with a physical therapist in a hospital setting, but unprepared for working in a remote area alone. Indeed, Spruit *et al*.[5] has noted that several PR programs may be run by only one healthcare professional and it is unknown whether

this healthcare professional has the appropriate clinical training to cover all aspects of a PR program. Moreover, a study by Johnston *et al.*[71] found that many PR healthcare professionals felt that they didn't have enough clinical training to operate a PR program in a rural or remote setting. These are possible reasons behind the unwarranted variation found in PR programs[4,5,38]. The QIs developed in this study are organized in the Donabedian framework and can identify specific aspects of PR programs that can improve participant health outcomes[40]. Through self-auditing, PR clinicians can potentially use these QIs as a checklist to ensure that their actions are adhering to the quality standards of PR.

2.10.4.3 Increasing Health Literacy

Finally, the list of QIs may improve adherence to a PR program because they break down the PR program into understandable steps (with an understandable process, measures, and desired outcomes) and if communicated to the patient, could improve health literacy for PR participants. One of the risk factors of low adherence in PR is low health literacy[72]. Health literacy is defined as the ability "*to obtain, process, and understand basic health-related information needed to make appropriate healthcare decisions*[72]." Health literacy refers to a patient's responsibility to understand relevant information for making proper health related decisions, but also a healthcare professional's and researcher's responsibility to make sure the information if comprehensible[72]. As PR is a complex intervention with many different facets, it may be difficult for a patient to understand the rationale behind different exercises, self-management techniques, and outcome measures. This lack of understanding of how PR works may make it difficult for a participant to adhere to the entire program. The final list of QIs breaks down the processes and outcomes of PR in a manner that is simple enough for a participant to understand. Future research should be done

to see whether participants can use QIs to improve their health literacy in PR, their self-efficacy, and their adherence.

2.10.5 Limitations

There are several limitations to this study. First, one of the methodological limitations is that the MRAM did not allow the participants to discuss the QIs that were rated appropriate in the first round. This is significant because during the discussion, many panelists realized that they may have misinterpreted some of the criteria definitions, particularly reliability and feasibility. This misinterpretation certainly could have applied for the QIs that were rated as appropriate. Another limitation of this study is that some of the suggestions made by the panel to change the wording of the discussed QIs could have equally applied to the QIs that were rated as appropriate. For example, based on the panel's suggestion, we modified QIs that had the word *"should"* since using this word made these QIs ambiguous and may have affected their reliability. However, in the QIs that were rated as appropriate, two QIs also used the word "*should.*" Therefore, before finalizing the list of QIs for implementation and operationalization, all of the QIs will be examined by the Canadian Thoracic Society (CTS) Chronic Obstructive Pulmonary Disease Clinical Assembly (research group overseeing this study). This group of respiratory experts will examine and modify the QIs for consistency and clarity in the wording.

There are also some disadvantages to the consensus methods in general. For example, given the same statements, different panel compositions may produce different ratings. Furthermore, familiarity with certain statements will lead to higher ratings for those statements, especially if there may be a lack of scientific evidence for a given statement but the panel may feel it is

important to have in clinical practice. It is also possible that repeating the study with a different panel would yield different results than those reported here. Different healthcare professionals from varying geographical regions, clinical experiences, and educational backgrounds will have different opinions regarding the QIs. Further, most of the evidence for the delivery of PR comes from COPD literature. It is possible that this study would have different results if the focus of PR was on non-COPD populations such as ILD and lung cancer. This study also included a PR participant in our panel, and it is recognized that participants in PR vary in their disease and experience of PR. That said, only nine QIs were discarded due to disagreement among the panel (DI > 1). Thus, we have confidence that our panel composition did not drastically affect the QIs chosen.

Another limitation of these QIs is that they are subject to change as research in PR improves. For example, in this study, QIs regarding interval training and inspiratory muscle training were rated as inappropriate and discarded, however these QIs may become important in the near future. Both of these interventions may be effective for a select cohort of PR participants, but not for the general PR population[30,73]. Additionally, as more research focuses on the delivery of PR to non-COPD PR populations, new and distinct interventions may emerge from the literature. In the future, these interventions may be become recommended by the ATS/ERS, and therefore the current QIs maybe require updating.

There are also aspects of PR that depend upon rules and regulations of a healthcare system. In this study, QIs that pertained to accessing patient data such as healthcare utilization, maintenance, and participant follow-up were rejected. During the panel discussion, many

panelists argued that it would not be feasible to include these QIs as it would difficult for auditors to access that type of data. These QIs were therefore rejected by the panel, indicating their expertise and understanding of how the delivery of PR works in Canada. From the results of this study, it is recommended that these QIs be routinely updated in order to reflect the current literature for best practice.

Chapter 3: Conclusion

In many developed countries it is estimated that up to 10% of the adult population has COPD[74]. The cumulative effects of the debilitating symptoms and comorbidities have a detrimental impact on a COPD patient's HRQoL and life expectancy[1]. The healthcare resources used by COPD patients create a great economic burden in many developed countries[1,6]. PR has been well-recognized as a key intervention for treating patients with COPD, and with proven success. However, many PR programs around the world differ in their delivery and organization thus threatening the quality of service[4,5,24,37]. Two PR jurisdictions[37,59] sought to rectify these issues through developing QIs, but these QI studies have several limitations. Thus, the purpose of this study was to develop QIs using MRAM, an established methodology. Through conducting a literature review, applying a standard framework, and conducting a formalized rating process, this study developed 56 QIs. While these QIs are similar in content to the two previous PR jurisdictions, they are written with more clarity and analyzed more thoroughly (with respect to importance, scientific soundness, reliability, and feasibility). While consensus has been established for these quality criteria for evaluating PR programs, further research is needed to pilot test the feasibility and reliability of implementing the QIs in different PR settings where resources vary. Such testing is crucial for the adoption of these QIs in clinical practice. Nonetheless, these 56 QIs could become the foundation for evaluating the quality of current PR programs across Canada and developing benchmarks for best practice.

References

1. Spruit MA, Singh SJ, Garvey C, ZuWallack R, Nici L, Rochester C, Hill K, Holland AE, Lareau SC, Man WD, Pitta F. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. American journal of respiratory and critical care medicine. 2013 Oct 15;188(8):e13-64.

2. Nici L, ZuWallack RL. Pulmonary rehabilitation: definition, concept, and history. Clinics in chest medicine. 2014 Jun 30;35(2):279-82.

3. Vestbo J, Hurd SS, Rodriguez-Roisin R. The 2011 revision of the global strategy for the diagnosis, management and prevention of COPD (GOLD)–why and what? The clinical respiratory journal. 2012 Oct 1;6(4):208-14.

4. Steiner MC, Roberts CM. Pulmonary rehabilitation: the next steps. The Lancet. Respiratory medicine. 2016 Mar;4(3):172.

5. Spruit MA, Pitta F, Garvey C, ZuWallack RL, Roberts CM, Collins EG, Goldstein R, McNamara R, Surpas P, Atsuyoshi K, López-Campos JL. Differences in content and organisational aspects of pulmonary rehabilitation programmes. European Respiratory Journal. 2014 May 1;43(5):1326-37.

6. Rochester CL, Vogiatzis I, Holland AE, Lareau SC, Marciniuk DD, Puhan MA, Spruit MA, Masefield S, Casaburi R, Clini EM, Crouch R. An official American Thoracic Society/European Respiratory Society policy statement: enhancing implementation, use, and delivery of pulmonary rehabilitation. American journal of respiratory and critical care medicine. 2015 Dec 1;192(11):1373-86.

7. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. The Cochrane Library. 2015 Feb.

8. Lacasse Y, Cates CJ, McCarthy B, Welsh EJ. This Cochrane Review is closed: deciding what constitutes enough research and where next for pulmonary rehabilitation in COPD.

9. Lee AL, Hill CJ, Cecins N, Jenkins S, McDonald CF, Burge AT, Rautela L, Stirling RG, Thompson PJ, Holland AE. The short and long term effects of exercise training in non-cystic fibrosis bronchiectasis–a randomised controlled trial. Respiratory research. 2014 Apr 15;15(1):44.

10. Morano MT, Araújo AS, Nascimento FB, da Silva GF, Mesquita R, Pinto JS, de Moraes Filho MO, Pereira ED. Preoperative pulmonary rehabilitation versus chest physical therapy in patients undergoing lung cancer resection: a pilot randomized controlled trial. Archives of physical medicine and rehabilitation. 2013 Jan 31;94(1):53-8.

11. Bajwah S, Ross JR, Peacock JL, Higginson IJ, Wells AU, Patel AS, Koffman J, Riley J. Interventions to improve symptoms and quality of life of patients with fibrotic interstitial lung disease: a systematic review of the literature. Thorax. 2013 Sep 1;68(9):867-79.

12. Ram FS, Robinson SM, Black PN. Effects of physical training in asthma: a systematic review. British Journal of Sports Medicine. 2000 Jun 1;34(3):162-7.

13. Wickerson L, Mathur S, Brooks D. Exercise training after lung transplantation: a systematic review. The Journal of Heart and Lung Transplantation. 2010 May 31;29(5):497-503.

14. Howley ET. Type of activity: resistance, aerobic and leisure versus occupational physical activity. Medicine & Science in Sports & Exercise. 2001 Jun 1;33(6):S364-9.

15.American College of Sports Medicine. ACSM's guidelines for exercise testing and prescription. Lippincott Williams & Wilkins; 2013 Mar 4.

16. Garvey C, Fullwood MD, Rigler J. Pulmonary rehabilitation exercise prescription in chronic obstructive lung disease: US survey and review of guidelines and clinical practices. Journal of cardiopulmonary rehabilitation and prevention. 2013 Sep 1;33(5):314-22.

17. O'Shea SD, Taylor NF, Paratz JD. Progressive resistance exercise improves muscle strength and may improve elements of performance of daily activities for people with COPD: a systematic review. Chest. 2009 Nov 30;136(5):1269-83.

18. Gloeckl R, Heinzelmann I, Baeuerle S, Damm E, Schwedhelm AL, Diril M, Buhrow D, Jerrentrup A, Kenn K. Effects of whole body vibration in patients with chronic obstructive pulmonary disease–a randomized controlled trial. Respiratory medicine. 2012 Jan 31;106(1):75-83.

19. Ng L, Chiang LK, Tang R, Siu C, Fung L, Lee A, Tam W. Effectiveness of incorporating Tai Chi in a pulmonary rehabilitation program for chronic obstructive pulmonary disease (COPD) in primary care—a pilot randomized controlled trial. European Journal of Integrative Medicine. 2014 Jun 30;6(3):248-58.

20. Magadle R, McConnell AK, Beckerman M, Weiner P. Inspiratory muscle training in pulmonary rehabilitation program in COPD patients. Respiratory medicine. 2007 Jul 31;101(7):1500-5.

21. Ries AL, Bauldoff GS, Carlin BW, Casaburi R, Emery CF, Mahler DA, Make B, Rochester CL, ZuWallack R, Herrerias C. Pulmonary rehabilitation: joint ACCP/AACVPR evidence-based clinical practice guidelines. Chest. 2007 May 31;131(5):4S-2S.

22. Bourbeau J, Julien M, Maltais F, Rouleau M, Beaupré A, Bégin R, Renzi P, Nault D, Borycki E, Schwartzman K, Singh R. Reduction of hospital utilization in patients with chronic

obstructive pulmonary disease: a disease-specific self-management intervention. Archives of internal medicine. 2003 Mar 10;163(5):585-91.

23. Spencer LM, Alison JA, McKeough ZJ. Maintaining benefits following pulmonary rehabilitation: a randomised controlled trial. European Respiratory Journal. 2010 Mar 1;35(3):571-7.

24. Camp PG, Hernandez P, Bourbeau J, Kirkham A, Debigare R, Stickland MK, Goodridge D, Marciniuk DD, Bhutani M, Dechman G. Pulmonary rehabilitation in Canada: a report from the Canadian thoracic society COPD clinical assembly. Canadian respiratory journal. 2015;22(3):147-52.

25. Rochester CL, Spanevello A. Heterogeneity of pulmonary rehabilitation: like apples and oranges–both healthy fruit. European Respiratory Journal. 2014 May 1;43(5):1223-6.

26. Johnston K, Grimmer-Somers K. Pulmonary rehabilitation: overwhelming evidence but lost in translation?. Physiotherapy Canada. 2010 Oct;62(4):368-73.

27. Jenkins S, Hill K, Cecins NM. State of the art: how to set up a pulmonary rehabilitation program. Respirology. 2010 Nov 1;15(8):1157-73.

28. Goodridge D, Marciniuk D. Rural and remote care: overcoming the challenges of distance. Chronic respiratory disease. 2016 May;13(2):192-203.

29. Gloeckl R, Marinov B, Pitta F. Practical recommendations for exercise training in patients with COPD.

30. Beauchamp MK, Nonoyama M, Goldstein RS, Hill K, Dolmage TE, Mathur S, Brooks D. Interval versus continuous training in individuals with chronic obstructive pulmonary disease-a systematic review. Thorax. 2010 Feb 1;65(2):157-64.

31. Puhan MA, Garcia-Aymerich J, Frey M, ter Riet G, Antó JM, Agustí AG, Gómez FP, Rodríguez-Roisín R, Moons KG, Kessels AG, Held U. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. The Lancet. 2009 Sep 4;374(9691):704-11.

32. Stickland MK, Jourdain T, Wong EY, Rodgers WM, Jendzjowsky NG, MacDonald GF. Using Telehealth technology to deliver pulmonary rehabilitation to patients with chronic obstructive pulmonary disease. Canadian respiratory journal. 2011;18(4):216-20.

33. Nield M, Hoo GW. Real-time telehealth for COPD self-management using Skype™. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2012 Nov 29;9(6):611-9.

34. Wennberg JE, Peters Jr PG. Unwarranted variations in the quality of health care: Can the law help medicine provide a remedy/remedies. Wake Forest L. Rev.. 2002;37:925.

35. Wennberg JE. Dealing with medical practice variations: a proposal for action. Health Affairs. 1984 May 1;3(2):6-32.

36. Mant J. Process versus outcome indicators in the assessment of quality of health care. International Journal for Quality in Health Care. 2001 Dec 1;13(6):475-80.

37. Yohannes AM, Stone RA, Lowe D, Pursey NA, Buckingham RJ, Roberts CM. Pulmonary rehabilitation in the United Kingdom. Chronic respiratory disease. 2011 Aug;8(3):193-9.

38. Dechman G, Hernandez P, Camp PG. Exercise prescription practices in pulmonary rehabilitation programs. Canadian Journal of Respiratory, Critical Care, and Sleep Medicine. 2017 Jun 6:1-7.

39. Steiner M, Holzhauer-Barrie J, Lowe D, Searle L, Skipper E, Welham S. Pulmonary Rehabilitation: Time to breathe better. National Chronic Obstructive Pulmonary Disease (COPD)

Audit Programme: Resources and organisation of Pulmonary Rehabilitation services in England and Wales. 2015.

40. Donabedian A. The quality of care: how can it be assessed?. Jama. 1988 Sep 23;260(12):1743-8.

41. Canadian Institute for Health Information. A Performance Measurement Framework for the Canadian Health System [Internet]. 2012. [cited 2017 July 22]. Available from: https://www.cihi.ca/en/hospital_framework_en.pdf

42. Kelley E, Hurst J. Health care quality indicators project.

43. Brooks D, Sottana R, Bell B, Hanna M, Laframboise L, Selvanayagarajah S, Goldstein R. Characterization of pulmonary rehabilitation programs in Canada in 2005. Canadian respiratory journal. 2007;14(2):87-92.

44. Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care. Qual Saf Health Care. 2002 Dec 1;11(4):358-64.

45. Gawande, A. The Checklist Manifesto. 2009. Atul Gawande, Better A Surg. Notes Perform. (2008).

46. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ: British Medical Journal (Online). 2016 May 3;353.

47. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AH, Dellinger EP, Herbosa T, Joseph S, Kibatala PL, Lapitan MC, Merry AF. A surgical safety checklist to reduce morbidity and mortality in a global population. New England Journal of Medicine. 2009 Jan 29;360(5):491-9.

48. Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis?. British Medical Journal. 2014 Jun 13;348:g3725.

49. Berwick DM. A user's manual for the IOM's 'Quality Chasm'report. Health affairs. 2002 May 1;21(3):80-90.

50. Donabedian A. Evaluating the quality of medical care. The Milbank Quarterly. 2005 Dec 1;83(4):691-729.

51. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ: British Medical Journal. 1995 Aug 5;311(7001):376.

52. Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, Marteau T. Consensus development methods: a review of best practice in creating clinical guidelines. Journal of health services research & policy. 1999 Oct;4(4):236-48.

53. To T, Guttmann A, Lougheed MD, Gershon AS, Dell SD, Stanbrook MB, Wang C, McLimont S, Vasilevska-Ristovska J, Crighton EJ, Fisman DN. Evidence-based performance indicators of primary care for asthma: a modified RAND Appropriateness Method. International Journal for Quality in Health Care. 2010 Dec 1;22(6):476-85.

54. Powell C. The Delphi technique: myths and realities. Journal of advanced nursing. 2003 Feb 1;41(4):376-82.

55. Nguyen HB, Corbett SW, Steele R, Banta J, Clark RT, Hayes SR, Edwards J, Cho TW, Wittlake WA. Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality. Critical care medicine. 2007 Apr 1;35(4):1105-12.

56. Zimmerman DR. Improving nursing home quality of care through outcomes data: the MDS quality indicators. International journal of geriatric psychiatry. 2003 Mar 1;18(3):250-7.

57. Linstone HA, Turoff M. The Delphi Method. Techniques and applications. 2002;53.

58. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR. The RAND/UCLA appropriateness method user's manual. RAND CORP SANTA MONICA CA; 2001.

59. Güell MR, Cejudo P, Rodríguez-Trigo G, Gàldiz JB, Casolive V, Regueiro M, Soler-Cataluña JJ. Standards for quality care in respiratory rehabilitation in patients with chronic pulmonary disease. Archivos de Bronconeumología (English Edition). 2012 Nov 30;48(11):396-404.

60. Marshall MN, Shekelle PG, McGlynn EA, Campbell S, Brook RH, Roland MO. Can health care quality indicators be transferred between countries?. Quality and Safety in Health Care. 2003 Feb 1;12(1):8-12.

61. Desveaux L, Janaudis-Ferreira T, Goldstein R, Brooks D. An international comparison of pulmonary rehabilitation: a systematic review. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2015 Mar 4;12(2):144-53.

62. FluidSurveys. FluidSurveys. [Internet] 2016. Available at: http://fluidsurveys.com/. (Accessed: 6th December 2016)

63. Esrailian E, Spiegel BM, Targownik LE, Dubinsky MC, Targan SR, Gralnek IM. Differences in the management of Crohn's disease among experts and community providers, based on a national survey of sample case vignettes. Alimentary pharmacology & therapeutics. 2007 Oct 1;26(7):1005-18.

64. Shekelle PG, MacLean CH, Morton SC, Wenger NS. Assessing care of vulnerable elders: methods for developing quality indicators. Annals of Internal Medicine. 2001 Oct 16;135(8_Part_2):647-52.

65. MacLean CH, Saag KG, Solomon DH, Morton SC, Sampsel S, Klippel JH. Measuring quality in arthritis care: methods for developing the Arthritis Foundation's quality indicator set. Arthritis Care & Research. 2004 Apr 15;51(2):193-202.

66. Uphoff EP, Wennekes L, Punt CJ, Grol RP, Wollersheim HC, Hermens RP, Ottevanger PB. Development of generic quality indicators for patient-centered cancer care by using a RAND modified Delphi method. Cancer nursing. 2012 Jan 1;35(1):29-37.

67. Van Engen-Verheul M, Kemps H, Kraaijenhagen RA, De Keizer N, Peek N. Modified Rand method to derive quality indicators: a case study in cardiac rehabilitation. InMIE 2011 Aug 15 (pp. 88-92).

68. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, Robinson N. Lost in knowledge translation: time for a map?. Journal of continuing education in the health professions. 2006 Dec 1;26(1):13-24.

69. Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. Journal of clinical epidemiology. 2005 Feb 28;58(2):107-12.

70. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. Annals of Internal medicine. 2005 Feb 15;142(4):260-73.

71. Johnston CL, Maxwell LJ, Alison JA. Establishing and delivering pulmonary rehabilitation in rural and remote settings: The opinions, attitudes and concerns of health care professionals. Australian Journal of Rural Health. 2016 Apr 1;24(2):106-14.

72. Omachi TA, Sarkar U, Yelin EH, Blanc PD, Katz PP. Lower health literacy is associated with poorer health status and outcomes in chronic obstructive pulmonary disease. Journal of general internal medicine. 2013 Jan 1;28(1):74-81.

73. Ambrosino N. The case for inspiratory muscle training in COPD.

74. Diaz-Guzman E, Mannino DM. Epidemiology and prevalence of chronic obstructive pulmonary disease. Clinics in chest medicine. 2014 Mar 31;35(1):7-16.

Appendices

Appendix A Final List of Quality Indicators

Quality Indicator and Rating Criteria Final Round	Median	Disagreement Index
Legend Blue = Appropriate Green = Discarded [Criteria is **bolded**] Appropriateness = Median Score 7-9 Discarded (Uncertain) = Median Score 4-6 Discarded (Disagreement Among Panel) = Disagreement Index > 1 Discarded (Inappropriate) = Median Score 1-3		
S1-Setting **The pulmonary rehabilitation program is held in a hospital, outpatient clinic, home or community setting**		
Importance	8	0.00
Scientific Soundness	8	0.20
Reliability	8	0.13
Feasibility	9	0.13
S2-Accessibility: Reworded Quality Indicator: The pulmonary rehabilitation program is located in a setting accessible by public and private transportation		
Importance	8	0.13
Scientific Soundness	8	0.30
Reliability	8	0.16
Feasibility	8	0.37
S3-Accessibility: Reworded: The pulmonary rehabilitation program is accessible to those with physical disabilities such as participants with visual impairments, obesity, and participants who require wheelchair access.		
Importance	8	0.43
Scientific Soundness	7	0.56
Reliability	7	0.37

Feasibility	7	0.37
S4-Space: Exercise Testing and Training **The pulmonary rehabilitation program has space for exercise training and testing**		
Importance	9	0.09
Scientific Soundness	9	0.13
Reliability	7	0.37
Feasibility	8	0.37
S5-Space: Education and Private Consultation **The pulmonary rehabilitation program has space for education and private consultation**		
Importance	8	0.13
Scientific Soundness	8	0.16
Reliability	7	0.56
Feasibility	8	0.16
S6-Aerobic Exercise Assessment and Monitoring Equipment **The pulmonary rehabilitation program monitors heart rate, oxygen saturation, blood pressure and dyspnea during aerobic exercise capacity testing**		
Importance	9	0.00
Scientific Soundness	9	0.13
Reliability	8	0.26
Feasibility	8	0.26
S7-Aerobic Exercise Assessment and Monitoring Equipment **The pulmonary rehabilitation program uses a cycle ergometer, or a treadmill or a flat open space to walk for aerobic exercise testing**		
Importance	8	0.09
Scientific Soundness	8	0.26
Reliability	8	0.30
Feasibility	8	0.11

Appendix A Final List of Quality Indicators

S8-Aerobic Exercise Training Equipment **The pulmonary rehabilitation program uses a cycle ergometer, treadmill, stairs a flat open space to walk for aerobic exercise training**		
Importance	8	0.00
Scientific Soundness	8	0.16
Reliability	7	0.37
Feasibility	8	0.16
S9-Aerobic Exercise Training and Monitoring Equipment **The pulmonary rehabilitation program that uses an arm ergometer for aerobic exercise training also uses a cycle ergometer or treadmill or stairs or a flat open space to walk.**		
Importance	7	0.65
Scientific Soundness	8	0.56
Reliability	8	0.00
Feasibility	7	0.62
S10-Aerobic Exercise Training and Monitoring Equipment **The pulmonary rehabilitation program should monitor heart rate, oxygen saturation, blood pressure and dyspnea during all exercise training**		
Importance	9	0.24
Scientific Soundness	8	0.09
Reliability	8	0.24
Feasibility	8	0.26
S11-Resistance Strength and Endurance Assessment Equipment **The pulmonary rehabilitation program uses any of the following equipment for upper and lower strength and endurance testing: portable dynamometer, weight machines, free weights, or computerized dynamometer.**		
Importance	8	0.30

Scientific Soundness	8	0.30
Reliability	7	0.52
Feasibility	8	0.00
S12-Resistance Strength and Endurance Equipment **The pulmonary rehabilitation program uses any of following equipment for upper and lower resistance training: weight machines, elastic bands, elastic tubing, or free weights**		
Importance	8	0.00
Scientific Soundness	8	0.00
Reliability	7	0.37
Feasibility	8	0.16
S13 Respiratory Muscle Assessment and Monitoring Equipment **The pulmonary rehabilitation program uses validated equipment for respiratory muscle assessment.**		
Importance	**3**	**0.52**
Scientific Soundness	7	0.52
Reliability	**5**	**0.65**
Feasibility	**5**	**0.97**
S14 Respiratory Muscle Training Equipment **The pulmonary rehabilitation program uses validated equipment for inspiratory muscle training**		
Importance	**4**	**0.52**
Scientific Soundness	**5**	**0.32**
Reliability	**7**	**1.04**
Feasibility	**5**	**0.97**
S15-Safety Equipment **The pulmonary rehabilitation program have an onsite defibrillator and staff trained in its use; or have access to advanced cardiac life support team (i.e. code blue team)**		
Importance	9	0.09

Scientific Soundness	8	0.09
Reliability	8	0.65
Feasibility	8	0.00
S16-Safety Equipment The pulmonary rehabilitation program has a safety/adverse event reporting system to the corresponding healthcare authority or healthcare institution		
Importance	8	0.09
Scientific Soundness	8	0.16
Reliability	7	0.65
Feasibility	8	0.30
S17-Safety Equipment The pulmonary rehabilitation program has supplemental oxygen equipment		
Importance	9	0.24
Scientific Soundness	9	0.24
Reliability	8	0.16
Feasibility	8	0.39
S18-Staff to Patient Ratio:Reworded Quality Indicator: During group exercise the patient to PR clinical staff ratio is no more than 8 to 1. For a complex case-mix the ratio is no more than 4 to 1; for PR participants with very severe disease the ratio is 1 to 1.		
Importance	8	0.30
Scientific Soundness	**6**	**0.85**
Reliability	**6**	**0.65**
Feasibility	8	0.30
S19-Reworded Quality Indicator: Program has an essential team of a physical therapist or respiratory therapist with access to a physician and/or a nurse.		
Importance	8	0.30

Scientific Soundness	**7**	**0.97**
Reliability	**7**	**0.16**
Feasibility	**7**	**0.16**
S20-Team of Essential Staff: Reworded Quality Indicator: Exercise testing and training is supervised by a physical therapist or a physician, or a kinesiologist/exercise physiologist with specialty training.		
Importance	8	0.26
Scientific Soundness	7	0.33
Reliability	**8**	**0.00**
Feasibility	7	0.15
S21-Additional Team Members The pulmonary rehabilitation programs has additional staff team members that may include psychologist, dietitian, pharmacist, or exercise specialist. This team of additional staff works in collaboration with the essential staff		
Importance	9	0.13
Scientific Soundness	8	0.39
Reliability	7	0.56
Feasibility	7	0.76
S22-Audit The pulmonary rehabilitation program has annual audits for quality assurance		
Importance	8	0.00
Scientific Soundness	7	0.72
Reliability	**8**	**0.16**
Feasibility	7	0.52
P1-Referral Participants are referred for a pulmonary rehabilitation program from healthcare professionals such as nurses, physicians and physiotherapists		
Importance	8	0.56

Scientific Soundness	6	0.52
Reliability	7	0.16
Feasibility	7	0.56
P2-Consent **Participants provide consent before participation in pulmonary rehabilitation**		
Importance	8	0.37
Scientific Soundness	8	0.64
Reliability	7	0.65
Feasibility	8	0.13
P3-Initial Clinical Evaluation **Participants have the required medical documents that confirm a respiratory diagnosis and indication for pulmonary rehabilitation**		
Importance	8	0.13
Scientific Soundness	8	0.00
Reliability	7	0.37
Feasibility	8	0.26
P4-Quality of Life Assessment Tests **Participants complete baseline and post-PR quality of life assessments**		
Importance	9	0.13
Scientific Soundness	8	0.20
Reliability	8	0.16
Feasibility	8	0.96
P5-Dyspnea Assessment Tests **Participants complete baseline and post-pulmonary rehabilitation dyspnea assessments**		
Importance	8	0.13
Scientific Soundness	8	0.26
Reliability	8	0.00
Feasibility	8	0.20

Appendix A Final List of Quality Indicators

P6-Fatigue Assessment Tests **Participants complete baseline and post-pulmonary rehabilitation fatigue assessments**		
Importance	7	0.37
Scientific Soundness	**6**	**0.65**
Reliability	8	0.16
Feasibility	7	0.37
P7-Symptom Assessment Tests **Participants complete baseline and post-pulmonary rehabilitation symptom assessments**		
Importance	9	0.37
Scientific Soundness	8	0.33
Reliability	8	0.16
Feasibility	8	0.30
P8-Physical Assessment Tests **Participants complete baseline and post-pulmonary rehabilitation assessments of the following: heart rate; oxygen saturation; and body mass index**		
Importance	8	0.37
Scientific Soundness	8	0.37
Reliability	8	0.37
Feasibility	8	0.24
P9-Functional Status Assessment Tests **Participants complete baseline and post-pulmonary rehabilitation functional status assessments**		
Importance	9	0.13
Scientific Soundness	8	0.26
Reliability	7	0.30
Feasibility	8	0.64
P10-Physical Activity Assessment Tests **Participants complete baseline and post-pulmonary rehabilitation physical activity assessments**		
Importance	9	0.13
Scientific Soundness	8	0.37

Reliability	8	0.16
Feasibility	8	0.16
P11-Psychological Assessment Tests **Participants complete baseline and post-pulmonary rehabilitation psychological assessments**		
Importance	8	0.56
Scientific Soundness	7	0.56
Reliability	8	0.16
Feasibility	7	0.16
P12-Self-Efficacy Tests **Participants complete baseline and post-pulmonary rehabilitation self-efficacy assessments**		
Importance	8	0.16
Scientific Soundness	7	0.33
Reliability	7	0.15
Feasibility	7	0.56
P13-Composite Outcomes Assessment Tests **Participants complete a baseline and post-pulmonary rehabilitation composite test**		
Importance	**5**	**0.32**
Scientific Soundness	**6**	**0.52**
Reliability	7	0.37
Feasibility	7	0.37
P14-Goal Setting **In consultation with pulmonary rehabilitation staff, participants develop written goals for the program**		
Importance	8	0.20
Scientific Soundness	7	0.52
Reliability	7	0.62
Feasibility	**6**	**0.58**
P15-Program Length **The length of the pulmonary rehabilitation program is at least 8 weeks**		

Importance	8	0.24
Scientific Soundness	8	0.24
Reliability	8	0.30
Feasibility	8	0.26
P16-Education: Reworded Quality Indicator: In addition to exercise training, the pulmonary rehabilitation program has an education and self-management component.		
Importance	8	0.13
Scientific Soundness	8	0.11
Reliability	8	0.16
Feasibility	7	0.56
P17-Education Topics The pulmonary rehabilitation program's education/self management component includes the following topics: anatomy, physiology and pathophysiology of lung disease; medication use; breathing techniques; anxiety and stress management; early recognition of exacerbation symptoms; dietary advice; energy conservation techniques; promotion of an active and healthy lifestyle; coping strategies; and secretion clearance techniques.		
Importance	9	0.13
Scientific Soundness	7	0.37
Reliability	7	0.33
Feasibility	7	0.56
P18-Warm-Up The pulmonary rehabilitation participants participate in a warm-up before each exercise session		
Importance	8	0.29
Scientific Soundness	7	0.37
Reliability	7	0.52
Feasibility	7	0.37

Appendix A Final List of Quality Indicators

P19-Cool-Down **The pulmonary rehabilitation participants engage in a cool down after each exercise session**		
Importance	7	0.46
Scientific Soundness	7	0.56
Reliability	6	1.04
Feasibility	5	1.04
P20-Aerobic Exercise Capacity Assessment Test Reworded: Participants undergo a baseline aerobic exercise capacity assessment using a Cardiopulmonary Exercise Test or an indirect measure (field walking test)		
Importance	8	0.33
Scientific Soundness	8	0.16
Reliability	8	0.37
Feasibility	8	0.16
P21-Reworded Quality Indicator: Participants receiving interval training have a program developed by the physical therapist, or a exercise physiologist, or a kinesiologist, or a physician.		
Importance	7	0.37
Scientific Soundness	8	0.37
Reliability	6	0.37
Feasibility	7	0.37
P22-Reworded Quality Indicator: An objective criterion is used to prescribe exercise intensity for continuous aerobic training.		
Importance	8	0.16
Scientific Soundness	7	0.16
Reliability	7	0.37
Feasibility	7	0.37
P23-Reworded Quality Indicator: The exercise prescription for interval aerobic training is based on an objective exercise test.		

Importance	8	0.00
Scientific Soundness	8	0.24
Reliability	6	0.37
Feasibility	7	0.65
P24-Aerobic Exercise Type One or more of the following forms of continuous and interval aerobic exercise training are used: stationary cycling, or treadmill walking, or walking, or stair climbing, or arm cycling (in conjunction with other aerobic exercise)		
Importance	8	0.13
Scientific Soundness	8	0.11
Reliability	8	0.30
Feasibility	8	0.30
P25-Exercise Progression Exercise intensity and duration should be increased as tolerated		
Importance	8	0.09
Scientific Soundness	8	0.00
Reliability	8	0.11
Feasibility	7	0.16
P26-Aerobic Exercise Frequency The pulmonary rehabilitation participants engage in aerobic exercise at least 3 times a week		
Importance	8	0.00
Scientific Soundness	8	0.00
Reliability	8	0.11
Feasibility	8	0.24
P27-Aerobic Exercise Duration: Reworded Quality Indicator: Exercise is structured to have participants complete at least 20 minutes of aerobic activity per session. Initially, shorter bouts of exercise may be used to accumulate 20 minutes of exercise.		
Importance	8	0.16

Scientific Soundness	7	0.52
Reliability	7	0.16
Feasibility	7	0.16
P28-Pulmonary rehabilitation Interval Training **PR program uses interval training that is safe but also produces a training effect**		
Importance	8	0.11
Scientific Soundness	7	0.30
Reliability	**6**	**0.97**
Feasibility	**5**	**0.97**
P29-Resistance Exercise Assessments for Muscle Strength **The pulmonary rehabilitation program participants are assessed for muscle strength**		
Importance	7	0.30
Scientific Soundness	7	0.26
Reliability	7	0.37
Feasibility	8	0.37
P30-Resistance Exercise Assessments for Muscle Endurance **The pulmonary rehabilitation program participants are assessed for muscle endurance**		
Importance	8	0.37
Scientific Soundness	7	0.37
Reliability	7	0.22
Feasibility	**6**	**0.65**
P31-Resistance Exercise Intensity Prescription Measure for Muscle Strength **Pulmonary rehabilitation participants are prescribed a strengthening exercise intensity based on their initial assessment**		
Importance	8	0.00
Scientific Soundness	7	0.30
Reliability	8	0.37

Feasibility	7	0.88
P32-Resistance Exercise Intensity: The pulmonary rehabilitation program participants start resistance training at 50%-70% of a direct or indirectly measured 1 Rep Max		
Importance	7	0.37
Scientific Soundness	**5**	**0.52**
Reliability	**6**	**0.97**
Feasibility	**6**	**0.97**
P33-Resistance Exercise Muscles to Target The pulmonary rehabilitation program resistance exercises target major upper limb muscles and lower limb muscles		
Importance	8	0.09
Scientific Soundness	8	0.09
Reliability	7	0.33
Feasibility	7	0.33
P34-Reworded Quality Indicator: Progression of resistance exercise is based on objective re-assessment of a participant's baseline exercise intensity and volume (repetitions and sets).		
Importance	8	0.00
Scientific Soundness	8	0.11
Reliability	**6**	**0.37**
Feasibility	7	0.37
P35-Resistance Exercise Frequency The pulmonary rehabilitation program participants participate in resistance exercise 2-3 times per week		
Importance	8	0.00
Scientific Soundness	8	0.00
Reliability	7	0.33
Feasibility	8	0.48

Appendix A Final List of Quality Indicators

P36 Resistance Repetitions and Sets **Resistance exercise includes at least 8 repetitions and 2-3 sets for each exercise**		
Importance	8	0.30
Scientific Soundness	7	0.33
Reliability	8	0.37
Feasibility	8	0.37
P37 Respiratory Muscle Assessments **Baseline respiratory muscle assessment is done according to a standardized test.**		
Importance	**4**	**0.52**
Scientific Soundness	7	0.65
Reliability	**6**	**0.32**
Feasibility	**6**	**0.97**
P38 Inspiratory Muscle Training Intensity **Participants begin their inspiratory respiratory muscle training at a PiMax of at least 30% and increase load as tolerated.**		
Importance	**4**	**0.65**
Scientific Soundness	**5**	**0.65**
Reliability	**6**	**0.97**
Feasibility	**5**	**1.04**
P39 Inspiratory Muscle Training Frequency **PR participants train their respiratory muscles at least 5 times a week, including home training**		
Importance	**5**	**0.65**
Scientific Soundness	7	0.56
Reliability	**6**	**0.97**
Feasibility	**6**	**0.97**
P40 Inspiratory Muscle Training Duration **Participants train their respiratory muscles at least 10-15 minutes daily**		
Importance	**5**	**0.65**

Scientific Soundness	**5**	**0.32**
Reliability	**4**	**0.85**
Feasibility	**4**	**0.32**
P41 Inspiratory Muscle Training Exercise Progression **The respiratory muscle training program is progressed as tolerated**		
Importance	**5**	**0.65**
Scientific Soundness	7	0.56
Reliability	**4**	**0.97**
Feasibility	**3**	**0.37**
P42 (Aerobic and Resistance Training) Flexibility Training **The pulmonary rehabilitation programs includes flexibility exercise training**		
Importance	8	0.56
Scientific Soundness	7	0.56
Reliability	**6**	**0.32**
Feasibility	**6**	**0.65**
P43-(Aerobic and Resistance Training) Supplemental Oxygen **PR program uses supplemental oxygen during exercise to maintain a safe oxygen saturation**		
Importance	8	0.00
Scientific Soundness	8	0.30
Reliability	8	0.56
Feasibility	8	0.65
P44-(Aerobic and Resistance Training) Emergency Response **The pulmonary rehabilitation program has a written response protocol for emergency situations**		
Importance	8	0.13
Scientific Soundness	8	0.11
Reliability	7	0.56
Feasibility	7	0.37

Appendix A Final List of Quality Indicators

P45-Reworded Quality Indicator: The pulmonary rehabilitation program has a maintenance program that provides access to exercise training and education.		
Importance	8	0.00
Scientific Soundness	8	0.24
Reliability	**6**	**0.65**
Feasibility	**6**	**0.65**
P46-Reworded Quality Indicator: The pulmonary rehabilitation program provides follow-up with participants after discharge to discuss issues related to the home exercise program and/or self-management.		
Importance	7	0.37
Scientific Soundness	**6**	**0.65**
Reliability	7	0.65
Feasibility	7	0.37
P47 (Aerobic and Resistance Training) Familiarization Participants are familiarized with exercise training by PR staff		
Importance	8	0.13
Scientific Soundness	8	0.30
Reliability	**6**	**0.65**
Feasibility	7	0.65
P48 (Aerobic and Resistance Training) Adherence PR program has a protocol for following up with participants that have missed a PR session and to encourage adherence		
Importance	8	0.30
Scientific Soundness	**5**	**0.00**
Reliability	**4**	**0.85**
Feasibility	**4**	**1.04**

Appendix A Final List of Quality Indicators

P49 Nutritional Supplementation **Eligible participants receive nutritional supplementation or referred to a dietitian**		
Importance	8	0.16
Scientific Soundness	**5**	**0.52**
Reliability	**6**	**0.37**
Feasibility	7	0.65
P50 Airway Clearance Techniques **Eligible participants receive teaching in airway clearance techniques**		
Importance	8	0.09
Scientific Soundness	8	0.48
Reliability	7	0.42
Feasibility	7	0.37
P51 Psychological support **Eligible participants receive psychological support or are referred to a psychologist**		
Importance	8	0.00
Scientific Soundness	7	0.30
Reliability	**6**	**0.52**
Feasibility	**6**	**0.65**
P52 Home Exercise Program **Reworded Quality Indicator: The pulmonary rehabilitation program provides exercise prescription and physical activity guidance for participants to exercise at home after discharge from pulmonary rehabilitation program.**		
Importance	8	0.09
Scientific Soundness	8	0.16
Reliability	7	0.37
Feasibility	7	0.16

Appendix A Final List of Quality Indicators

P53-Health Care Utilization **The pulmonary rehabilitation program programs assess healthcare utilization before and after pulmonary rehabilitation, or if hospitalization occurs during the pulmonary rehabilitation program**		
Importance	7	0.56
Scientific Soundness	**6**	**0.32**
Reliability	**5**	**0.85**
Feasibility	**4**	**0.32**
O1-Quality of Life Outcome Measures **Quality of life outcomes are measured after the PR program**		
Importance	8	0.11
Scientific Soundness	8	0.09
Reliability	7	0.46
Feasibility	8	0.39
O2 -Dyspnea Outcome Measures **Dyspnea outcomes are measured after the PR program**		
Importance	8	0.11
Scientific Soundness	8	0.13
Reliability	7	0.39
Feasibility	8	0.26
O3-Reworded Quality Indicator: Fatigue is assessed before and after the pulmonary rehabilitation program.		
Importance	7	0.62
Scientific Soundness	**5**	**0.65**
Reliability	8	0.37
Feasibility	8	0.37
O4 -Symptom Outcome Measures **Symptom outcomes are assessed after the PR program**		
Importance	8	0.11
Scientific Soundness	8	0.11

Reliability	8	0.30
Feasibility	8	0.11
O5-Physical Outcome Measures Physical outcomes are measured after rehabilitation with one or more of the following tests: heart rate; oxygen saturation; body mass index		
Importance	8	0.30
Scientific Soundness	8	0.16
Reliability	7	0.62
Feasibility	7	0.26
O6-Functional Outcome Measures Functional performance outcomes are assessed after the PR program		
Importance	8	0.30
Scientific Soundness	8	0.11
Reliability	7	0.37
Feasibility	8	0.37
O7-Reworded Quality Indicator: Physical activity is assessed before and after the pulmonary rehabilitation program.		
Importance	8	0.16
Scientific Soundness	8	0.30
Reliability	7	0.37
Feasibility	**6**	**0.37**
O8-Reworded Quality Indicator: An objective psychological status outcome is assessed before and after the PR program.		
Importance	7	0.26
Scientific Soundness	7	0.37
Reliability	8	0.16
Feasibility	8	0.37

Appendix A Final List of Quality Indicators

O9-Reworded Quality Indicator: Healthcare utilization is assessed before and after the pulmonary rehabilitation program.		
Importance	7	0.44
Scientific Soundness	7	0.56
Reliability	**6**	**0.65**
Feasibility	**5**	**1.04**
O10 -Self-efficacy Outcome Measures Self-efficacy outcomes are measured assessed after the PR program		
Importance	8	0.16
Scientific Soundness	8	0.16
Reliability	7	0.52
Feasibility	7	0.26
O11- Reworded Quality Indicator: An objective composite outcome is assessed before and after the pulmonary rehabilitation program.		
Importance	**5**	**0.85**
Scientific Soundness	7	0.62
Reliability	**6**	**0.32**
Feasibility	**6**	**0.65**
O12-Aerobic Exercise Capacity Outcome Measures Exercise capacity outcomes are assessed after the PR program		
Importance	8	0.00
Scientific Soundness	8	0.00
Reliability	8	0.30
Feasibility	8	0.69
O13-Reworded Quality Indicator: An objective measure is used to assess strength before and after the pulmonary rehabilitation program.		
Importance	8	0.16

Scientific Soundness	7	0.16
Reliability	**7**	**1.04**
Feasibility	**7**	**1.04**
O14-Reworded Quality Indicator: An objective measure is used to assess muscle endurance before and after the pulmonary rehabilitation program.		
Importance	7	0.16
Scientific Soundness	7	0.30
Reliability	**6**	**0.37**
Feasibility	**6**	**0.65**
O15-Reworded Quality Indicator: An objective respiratory muscle strength outcome is completed before and after the pulmonary rehabilitation program.		
Importance	**5**	**0.32**
Scientific Soundness	**5**	**0.97**
Reliability	**7**	**1.04**

Appendix B IPRAS Calculation

IRPAS = 2.35+(1.5*[Asymnmetry index])

The constant 2.35 describes the IPR if perfect symmetry exists. This was determined by RAND/UCLA organization in previous experiments on large datasets. 1.5 is a constant and the correction for asymmetry in the dataset. The asymmetry index (AI) is defined as "*the distance between the central point IPR and the central point of the 1-9 [Likert] scale*[58]." The AI therefore differs for every QI rated[58]. AI is calculated according to the following formula AI= Abs(5-[central point IPR]), where central point IPR = ([70th percentile]+[30th percentile])/2

Appendix C Microsoft Excel Equations for Calculating Appropriateness, Inappropriateness, and the Disagreement Index

1. Median: median (x, x)

2. 30^{th} percentile: (PERCENTILE.EXC(x:x;0,3)

3. 70^{th} percentile: (PERCENTILE.EXC(x:x;0,7)

4. Interpercentile Range of the 30^{th}-70^{th} Percentiles: (=[70th percentile] – [30th percentile])

5. Central Point Interpercentile Range: ((=[70th percentile] + [30th percentile])/2)

6. Asymmetry Index: (=ABS(5-[Central Point Interpercentile Range])

7. Interpercentile Range Adjusted for Symmetry: (=2.35 + (1.5*[Asymmetry Index])

8. Disagreement Index: (Interpercentile Range of the 30^{th}-70^{th} Percentiles/ Interpercentile Range Adjusted for Symmetry)

9. QI Criterion Score is rated Appropriate: (=IF(Median>=7,TRUE))

10. QI Criterion Score is rated as Uncertain: (=IF((AND(Median>=4, Median<=6)), TRUE, FALSE)

11. QI Criterion Score is rated as Inappropriate (=IF(Median<=3,TRUE))

12. QI Criterion is rated as having disagreement among the panel: (=Disagreement Index>1, TRUE)

Appendix D Reworded Quality Indicators from Panel Discussion

Appendix D Reworded Quality Indicators from Panel Discussion

Quality Indicator and Rating Criteria	Quality Indicator and Rating Criteria	Median	Disagreement Index
Legend Yellow = Discarded Appropriateness = Median Score 7-9 Uncertain = Median Score 4-6 [Criteria is **bolded**] Disagreement Index > 1 [Criteria is **bolded**] Inappropriate = Median Score 1-3 [Criteria is **bolded**]			
S2-Accessibility **The pulmonary rehabilitation program is located in a setting reasonably accessible by public and private transportation.**	**S2-Accessibility: Reworded Quality Indicator: The pulmonary rehabilitation program is located in a setting accessible by public and private transportation**		
Importance	Importance	8	0.13
Scientific Soundness	Scientific Soundness	8	0.30
Reliability	Reliability	8	0.16
Feasibility	Feasibility	8	0.37
S3-Accessibility **The pulmonary rehabilitation program is reasonably accessible to those with physical disabilities such as participants with visual impairments and participants who require wheelchair access**	**S3-Accessibility:** **Reworded: The pulmonary rehabilitation program is accessible to those with physical disabilities such as participants with visual impairments, obesity, and participants who require wheelchair access.**		
Importance	Importance	8	0.43
Scientific Soundness	Scientific Soundness	7	0.56
Reliability	Reliability	7	0.37
Feasibility	Feasibility	7	0.37

Appendix D Reworded Quality Indicators from Panel Discussion

Quality Indicator and Rating Criteria	Quality Indicator and Rating Criteria	Median	Disagreement Index
S18-Staff to Patient Ratio During group exercise the patient to pulmonary rehabilitation staff ratio is no more than 8 to 1	**S18-Staff to Patient Ratio: Reworded Quality Indicator: During group exercise the patient to PR clinical staff ratio is no more than 8 to 1. For a complex case-mix the ratio is no more than 4 to 1; for PR participants with very severe disease the ratio is 1 to 1.**		
Importance	Importance	8	0.30
Scientific Soundness	**Scientific Soundness**	**6**	**0.85**
Reliability	**Reliability**	**6**	**0.65**
Feasibility	Feasibility	8	0.30
S20-Team of Essential Staff Exercise testing and training should be supervised by a physical therapist or a physician who have expertise in testing and training	**S20-Team of Essential Staff: Reworded Quality Indicator: Exercise testing and training is supervised by a physical therapist or a physician, or a kinesiologist/exercise physiologist with specialty training.**		
Importance	Importance	8	0.26
Scientific Soundness	Scientific Soundness	7	0.33
Reliability	Reliability	8	0.00
Feasibility	Feasibility	7	0.15
P16-Education In addition to exercise training, the pulmonary rehabilitation program has a comprehensive education/self-management component	**P16-Education: Reworded Quality Indicator: In addition to exercise training, the pulmonary rehabilitation program has an education and self-management component.**		
Importance	Importance	8	0.13
Scientific Soundness	Scientific Soundness	8	0.11

Appendix D Reworded Quality Indicators from Panel Discussion

Quality Indicator and Rating Criteria	Quality Indicator and Rating Criteria	Median	Disagreement Index
Reliability	Reliability	8	0.16
Feasibility	Feasibility	7	0.56
P20-Aerobic Exercise Capacity Assessment Test **Participants undergo a baseline pulmonary rehabilitation aerobic exercise capacity assessment**	**P20-Aerobic Exercise Capacity Assessment Test Reworded: Participants undergo a baseline aerobic exercise capacity assessment using a Cardiopulmonary Exercise Test or an indirect measure (field walking test)**		
Importance	Importance	8	0.33
Scientific Soundness	Scientific Soundness	8	0.16
Reliability	Reliability	8	0.37
Feasibility	Feasibility	8	0.16
P21-Indication for Interval Training **Indication for Interval Training: Participants that require interval training have a program developed by the physical therapist or physician**	**P21-Reworded Quality Indicator: Participants receiving interval training have a program developed by the physical therapist, or an exercise physiologist, or a kinesiologist, or a physician.**		
Importance	Importance	7	0.37
Scientific Soundness	Scientific Soundness	8	0.37
Reliability	**Reliability**	6	0.37
Feasibility	Feasibility	7	0.37
P22-Continuous Aerobic Exercise Intensity **Participants should work at a continuous aerobic intensity that produces a training effect**	**P22-Reworded Quality Indicator: An objective criterion is used to prescribe exercise intensity for continuous aerobic training.**		
Importance	Importance	8	0.16
Scientific Soundness	Scientific Soundness	7	0.16

Appendix D Reworded Quality Indicators from Panel Discussion

Quality Indicator and Rating Criteria	Quality Indicator and Rating Criteria	Median	Disagreement Index
Reliability	Reliability	7	0.37
Feasibility	Feasibility	7	0.37
P27-Aerobic Exercise Frequency and Duration: **Continuous aerobic exercise training should be at least 20 min per exercise session**	**P27-Aerobic Exercise Duration: Reworded Quality Indicator: Exercise is structured to have participants complete at least 20 minutes of aerobic activity per session. Initially, shorter bouts of exercise may be used to accumulate 20 minutes of exercise.**		
Importance	Importance	8	0.16
Scientific Soundness	Scientific Soundness	7	0.52
Reliability	Reliability	7	0.16
Feasibility	Feasibility	7	0.16
P29-Resistance Exercise Assessments for Muscle Strength **The pulmonary rehabilitation program participants are assessed for muscle strength**	**P29-Resistance Exercise Assessments for Muscle Strength** **The pulmonary rehabilitation program participants are assessed for muscle strength**		
Importance	Importance	7	0.30
Scientific Soundness	Scientific Soundness	7	0.26
Reliability	Reliability	7	0.37
Feasibility	Feasibility	8	0.37
P45 (Aerobic and Resistance Training) Maintenance and Follow-up **The PR program has a maintenance program that provides continued exercise training and education for the participant**	**P45-Reworded Quality Indicator: The pulmonary rehabilitation program has a maintenance program that provides access to exercise training and education.**		
Importance	Importance	8	0.00

Appendix D Reworded Quality Indicators from Panel Discussion

Quality Indicator and Rating Criteria	Quality Indicator and Rating Criteria	Median	Disagreement Index
Scientific Soundness	Scientific Soundness	8	0.24
Reliability	**Reliability**	**6**	**0.65**
Feasibility	**Feasibility**	**6**	**0.65**
P46 (Aerobic and Resistance Training) Maintenance and Follow Up **Program follows up patient after discharge via telephone to discuss participant's exercise program and answer questions**	**P46-Reworded Quality Indicator: The pulmonary rehabilitation program provides follow-up with participants after discharge to discuss issues related to the home exercise program and/or self-management.**		
Importance	**Importance**	7	0.37
Scientific Soundness	**Scientific Soundness**	**6**	**0.65**
Reliability	**Reliability**	7	0.65
Feasibility	**Feasibility**	7	0.37
P52 Home Exercise Program **The pulmonary rehabilitation program provides exercise prescription or physical activity guidance for participants to exercise at home**	**P52 Home Exercise Program Reworded Quality Indicator: The pulmonary rehabilitation program provides exercise prescription and physical activity guidance for participants to exercise at home after discharge from pulmonary rehabilitation program.**		
Importance	Importance	8	0.09
Scientific Soundness	Scientific Soundness	8	0.16
Reliability	Reliability	7	0.37
Feasibility	Feasibility	7	0.16
O3-Fatigue Outcome Measures **Fatigue outcomes are assessed after the PR program.**	**O3-Reworded Quality Indicator: Fatigue is assessed before and after the pulmonary rehabilitation program.**		

Appendix D Reworded Quality Indicators from Panel Discussion

Quality Indicator and Rating Criteria	Quality Indicator and Rating Criteria	Median	Disagreement Index
Importance	Importance	7	0.62
Scientific Soundness	**Scientific Soundness**	**5**	**0.65**
Reliability	Reliability	8	0.37
Feasibility	Feasibility	8	0.37
O7-Physical Activity Outcome Measures **Physical activity outcomes are assessed after the PR program**	**O7-Reworded Quality Indicator: Physical activity is assessed before and after the pulmonary rehabilitation program.**		
Importance	Importance	8	0.16
Scientific Soundness	Scientific Soundness	8	0.30
Reliability	Reliability	7	0.37
Feasibility	**Feasibility**	**6**	**0.37**
O8-Psychological Outcome Measures **Psychological outcomes are assessed after the PR program.**	**O8-Reworded Quality Indicator: An objective psychological status outcome is assessed before and after the PR program.**		
Importance	Importance	7	0.26
Scientific Soundness	Scientific Soundness	7	0.37
Reliability	Reliability	8	0.16
Feasibility	Feasibility	8	0.37
O9-Health Care Utilization Outcome Measures **Healthcare utilization outcomes are assessed after the PR program**	**O9-Reworded Quality Indicator: Healthcare utilization is assessed before and after the pulmonary rehabilitation program.**		
Importance	Importance	7	0.44
Scientific Soundness	Scientific Soundness	7	0.56
Reliability	**Reliability**	**6**	**0.65**
Feasibility	**Feasibility**	**5**	**1.04**

Appendix D Reworded Quality Indicators from Panel Discussion

Quality Indicator and Rating Criteria	Quality Indicator and Rating Criteria	Median	Disagreement Index
O9-Health Care Utilization Outcome Measures Healthcare utilization outcomes are assessed after the PR program	**O9-Reworded Quality Indicator: Healthcare utilization is assessed before and after the pulmonary rehabilitation program.**		
Importance	Importance	7	0.44
Scientific Soundness	Scientific Soundness	7	0.56
Reliability	**Reliability**	**6**	**0.65**
Feasibility	**Feasibility**	**5**	**1.04**
O11- Composite Outcome Measures Composite outcomes are assessed after the PR program	**O11- Reworded Quality Indicator: An objective composite outcome is assessed before and after the pulmonary rehabilitation program.**		
Importance	**Importance**	**5**	**0.85**
Scientific Soundness	Scientific Soundness	7	0.62
Reliability	**Reliability**	**6**	**0.32**
Feasibility	**Feasibility**	**6**	**0.65**
O13-Muscle Strength Outcome Measures Muscle strength outcomes are assessed after the PR program	**O13-Reworded Quality Indicator: An objective measure is used to assess strength before and after the pulmonary rehabilitation program.**		
Importance	Importance	8	0.16
Scientific Soundness	Scientific Soundness	7	0.16
Reliability	**Reliability**	**7**	**1.04**
Feasibility	**Feasibility**	**7**	**1.04**

Appendix D Reworded Quality Indicators from Panel Discussion

Quality Indicator and Rating Criteria	Quality Indicator and Rating Criteria	Median	Disagreement Index
O14-Muscle Endurance Outcomes Measures **Muscle endurance outcomes are assessed after the PR program**	**O14-Reworded Quality Indicator: An objective measure is used to assess muscle endurance before and after the pulmonary rehabilitation program.**		
Importance	Importance	7	0.16
Scientific Soundness	Scientific Soundness	7	0.30
Reliability	**Reliability**	**6**	**0.37**
Feasibility	**Feasibility**	**6**	**0.65**
O15-Respiratory Muscle Strength Outcome Measures **Respiratory muscle strength outcomes are assessed after the PR program**	**O15-Reworded Quality Indicator: An objective respiratory muscle strength outcome is completed before and after the pulmonary rehabilitation program.**		
Importance	**Importance**	**5**	**0.32**
Scientific Soundness	**Scientific Soundness**	**5**	**0.97**
Reliability	**Reliability**	**7**	**1.04**
Feasibility	**Feasibility**	**5**	**0.97**