SCREENING INTERVALS AND THE RISK OF CARCINOMA *IN SITU* OF THE CERVIX

by

NORMAN PHILLIPS

B.A., McGill University, 1978
M.A., The University of British Columbia, 1984

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Statistics)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1994

© Norman Phillips, 1994

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)_____

Department of Statistics

The University of British Columbia
Vancouver, Canada

Date Sept. 6, 1994

Abstract

This study examines the effect of length of interval between routine tests on the risk of carcinoma *in situ* (CIS) of the uterine cervix using cohort data from the B.C. population. CIS is a symptomless disease only detected by screening. Because of this, special methods are required for estimating incidence rates. Some case-control studies have used prevalence odds ratios to estimate the relative risk of disease, usually invasive cancer, from length of screening interval. But duration of disease is related to interval length and hence prevalence rates cannot be used to estimate relative risk. A multivariate model is fit to the incidence data using Poisson regression, and prevalence rates are fit with a logistic regression model. The results for prevalence odds ratios indicate a positive association between screening interval length and risk of disease whereas the results for relative risk indicate a negative relationship. Theoretical screening models are considered to examine the consequences of a case-control paradigm in which controls are matched with cases on the basis of having had a screen near the date of diagnosis of the case, the matching period. As the matching period shortens, the distribution of interval lengths for controls converges to the underlying distribution, whereas the distribution of interval lengths for cases equals the distribution of lengths of intervals which span a point in time. The latter distribution favours longer intervals. The difference is not due to the sampling of controls but, rather, to the relation between interval length and duration of disease. A matched case-control study is simulated with the cohort data, and a

conditional likelihood logistic regression model is fit. The results agree with those of a logistic regression analysis of prevalence rates indicating a positive relation between interval length and risk of disease. When the sampling of controls is weighted by interval length the odds ratios approximate the relative risk. A possible explanation of the surprising result that screening interval length is inversely related to risk of diagnosis of CIS is that more cases are cured with time by the natural regression of disease than by treatment of earlier stages of disease. On the other hand, incidence rate is negatively related to recency and frequency of prior negative screens, possibly because of the occurrence of false negative tests. However, the effect of regression predominates and the unavoidable conclusion is that less frequent screening decreases the risk of diagnosis of CIS.

Table of Contents

List of Tables

List of Figures

Acknowledgement

I would like to thank Andy Coldman for all the valuable time and excellent supervision he has provided, Nhu Le for helping me out in a "crunch", Brenda Morrison for her willingness to offer advice and assistance at all times, the B.C. Cancer Agency for supplying the data, and finally, my wife, Gloria, for taking on my share of the childcare, laundry, and dishes while I finished this project.

Chapter 1

Introduction

Cervical cancer, i.e. cancer of the uterine cervix, appears to develop in stages. "There is a continuous spectrum of histological [cell] abnormalities of the squamous epithelium [type of cell] of the cervix [the opening of the uterus] from dysplasia through carcinoma *in situ* (CIS) to micro-invasive and invasive lesions [cancer]"[1]. It is generally believed that every case of invasive cancer of the cervix originated in dysplasia and then progressed to carcinoma *in situ* before developing into cervical cancer. Not all cases of dysplasia develop into invasive cancer, and the time to cancer development in those which do progress is variable.

It is generally believed that the earlier cancer is detected and treated, the better the prognosis. Since the stages of dysplasia and CIS are non-symptomatic they are only discovered by accident or by screening. Screening is the practice of testing for the presence of disease in the absence of outward signs of disease. The rationale behind screening is to enable the treatment of the disease at a stage when prognosis is better and treatment may be less invasive. The Papanicolaou (PAP) test is a procedure for detecting cell abnormalities which are believed to be precursors of cervical cancer and is the screening method of choice for cervical cancer. A representative cell sample is obtained

using an Ayre's spatula from the transformation zone of the uterine cervix. The sample is placed on a slide to be examined by microscope. Test results rate the degree of cell abnormality on a scale from one to four, with "class" ones being normal. Abnormal results may be followed-up with more PAP tests or the individual may be referred for further examination.

Prior to the introduction of colposcopy in the mid 1970s, treatment generally consisted of cone biopsy or hysterectomy. Both procedures are relatively serious and so treatment was not recommended unless there was evidence of persistent abnormality or the abnormality was sufficiently serious. Colposcopy[2] is a method of visually inspecting the cervix. This allows the localization of the treatment with such methods as cryotherapy (freezing) and laser therapy. Since these methods are relatively non-aversive and have fewer complications, they are applied quite soon after signs of abnormality.

The effectiveness of the PAP test as a screen for cervical cancer is unknown since randomized trials have never been carried out. Screening was implemented before its effectiveness was known[3]. The indirect evidence is sufficiently compelling as to render future randomized controlled trials unethical. "The most persuasive evidence that screening for cervical cancer is effective comes from comparisons of cervical cancer in populations which introduced mass screening with different intensities and at different times"[3]. For example, in B.C., the incidence of clinically invasive cancer and associated morbidity have decreased by about 75% since the introduction of cervical screening[4].

However, inferences must be drawn from observational studies which have various potential sources of bias[5]. Parkins et. al.[6] report risk factors such as SES (social-economic status) to be related to the probability of screening. However, adjusting for SES does not seem to materially affect the relation between screening patterns and risk of cervical neoplasia[7]. Knox[8] argues that case-control studies are invalid if the factors which predispose to disease also affect the likelihood of screening. The same criticism also applies to cohort studies. In spite of the inherent flaws of observational studies, the significant reduction in deaths from cervical cancer has been attributed to screening using the PAP test[9,3].

One of the risks associated with screening is that individuals might undergo unnecessary treatment which in itself carries some risk. There is some evidence that stages even as far along as carcinoma *in situ* (CIS) have significant rates of regression, i.e., return to normal spontaneously, especially among younger individuals[1]. Since it is considered unethical to withhold treatment from patients with cell abnormalities, it is impossible to estimate regression and progression rates directly[10], except in groups which decline treatment. So indirect methods have been employed to try to estimate regression rates. One method which has been employed for this purpose compares the estimates of incidence (the rate of new cases in a time interval) and prevalence (the proportion of population who are cases at a given point in time) at different stages of disease on the assumption that if all cases of disease progress, then current prevalence cases at one stage would eventually become incidence cases at the next[1]. We should find that,

> Prevalence of carcinoma *in situ* or worse at time t_1+Total incidence of carcinoma *in situ* or worse during the interval t_1-t_2 = Prevalence of carcinoma *in situ* or worse at time t_2+Total incidence of clinical cancer during interval t_1-t_2[1].

Typically such calculations indicate a lower than expected number of cases of more severe disease. Such shortfalls are attributed to regression. But the results could also be due to false negative test results, i.e. results which indicate absence of disease when, in fact, disease is present. Incidence rates of earlier stages of disease may be inflated and the prevalence rates of the same may be deflated as a result of false negative test results[1]. False negatives, along with inadequate screening, have been blamed for some of the fatalities observed[11]. The may arise from "inadequate cell collection, smear preparations or smear interpretation"[11]. Estimates of false negative rates vary from 6% to 55%[11].

Current research attempts to model the risk of disease, generally invasive cancer, as a function of screening patterns using data from observational studies and keeping in mind the potential impact of disease regression and false negative test results. Screen-detected diseases require special methods for estimating incidence rates because the time of onset is unknown. Although multivariate models are not typically employed with incidence rate data, they are well suited for Poisson regression models. Case-control methods have been used to approximate risk of disease by prevalence odds ratios. But this is not appropriate when the exposure variable is related to duration of disease. Some case-controls studies select controls on the basis of screening times, which would seem to introduce a dependency between exposure variable status and sampling probabilities.

This issue is explored with theoretical screening models. The distributions of interval lengths for cases and controls are derived under the null hypothesis that there is no relation between interval length and risk of disease. Finally these issues will be illustrated with analyses using data from the B.C. screening program.

Chapter 2

Review of Literature

D.A. Boyes, B. Morrison, and colleagues[1], undertook a cohort study with data from the British Columbia Cervical Cytology Screening Program covering the years 1949-1969. The data consisted of two cohorts of individuals who were born between 1914 and 1918 or between 1929 and 1933 and who had been screened at least once prior to Jan. 1, 1970. The objective of the study was to provide estimates of prevalence and incidence rates of dysplasia or worse and carcinoma *in situ* or worse. Incidence rates of carcinoma *in situ* or worse were estimated from the number of cases developing while under surveillance relative to the total accumulated time at risk in the sample. Prevalence rates were estimated from "abnormalities discovered at first contact". An alternative estimate of prevalence is also given by the proportion of the population who are cases at any given point in time. These estimates took into account losses due to death and population variation due to immigration and emigration. The initial estimates of cumulative incidence of carcinoma *in situ* or worse were 17.9 per 1000 for ages 18 to 38 (Cohort 2), 19.3 per 1000 for ages 33 to 53 (Cohort 1) and 29.8 per 1000 for ages 18 to 53 (both cohorts).

Being an observational study, there were concerns about the possibility of sample biases. In the beginning years of the program, testing was done primarily as a support service for the diagnosis of symptomatic women. Thus the estimates of incidence and prevalence tend to be inflated. There is a tendency for higher SES females to be screened more readily although an effort was made to recruit women with low SES backgrounds. It is unclear what effect this bias might have on the estimates. Third, those who enter the program at older ages are different from those who enter at younger ages. A variety of factors may be involved in this selection bias.

Boyes et. al. examined two potential sources of false negatives, namely errors in the lab and sampling (literally) errors. The rate of lab errors was determined to be about 8.5% for Cohort 1 and 15.7% for Cohort 2. These estimates were based on review of tests which were originally coded negative but the patient subsequently developed the disease and interpolated to women who only presented once for examination. This translates into a dating error of about 26-38.6 months based on the average interval between the date of the actual previous negative and the original date of diagnosis.

Boyes et. al. refer to the remaining false negatives as "residual" false negatives and attempt to infer their rate from "the difference between apparent incidence rates derived from short intervals between examinations and the rates derived from long intervals between examinations". They also considered the patients history of tests on the assumption that "after several smears have been taken any positive case is likely to be

a genuinely new case since, although a case may have been missed at one examination, it is unlikely to have been missed at two or three successive examinations". Incidence rates were calculated for various combinations of interval length and numbers of previous negative test. The authors observed that short intervals with few preceding negative tests tend to have the highest rates. So they based their estimates of rates on those obtained from long intervals with many preceding negative tests.

> The rationale of the method is straightforward. First, cases discovered after a short interval from a previous smear are likely to be due to a classification error on the earlier smear. Secondly, after several smears have been taken any positive case is likely to be a genuinely new case since, although a case may have been missed at examination, it is unlikely to have been missed at 2 or 3 successive examinations.

Table 1 presents the corrected incidence rate estimates based on Table 18 in Boyes et al. Using these corrected incidence rates, prevalence estimates were revised to adjust for cases "who were not new 'incidence' cases but were missed 'prevalence' cases"[1]

The discrepancy between observed and corrected incidence rates provided an estimate of the false negative rate for various age groups. These ranged from 4.1% in the 45-49 age group to 12.9% in the 25-29 age group. Finally, the data were retabulated, correcting for estimated false negatives and adjusting the date of onset for such cases by the amount observed for lab false negatives.

Next, Boyes et. al. compared prevalence and incidence estimates in an effort to estimate

Table 1. Corrected incidence rate of CIS or worse per 1000 women years

Age range	Cohort	Incidence rate
25-29	2	0.98
30-34	2	1.04
35-39	2	0.36
40-44	1	0.46
45-49	1	0.60
50-54	1	0.32

the rate of disease regression. They reasoned that

> if pre-clinical cancer progresses to clinical cancer, then at the end of any time period the accumulated incidence of pre-clinical cancer should be equal to the prevalence of pre-clinical cancer plus the accumulated incidence of clinical cancer. If all the possible sources of error ... have been taken into account, a ratio of prevalence of pre-clinical cancer plus accumulated incidence of clinical cancer to accumulated incidence of pre-clinical cancer substantially lower than unity must imply an excess incidence of pre-clinical cancer and indicate that regression is part of the natural history.[1]

Combining the two cohorts, Boyes et. al. estimate this ratio to be 0.59-0.96 for carcinoma *in situ* or worse. Thus there is some evidence that regression occurs.

The method of estimating incidence rates for varying numbers of previous negative tests and for varying intervals since the last negative tests has become a research paradigm. Several studies utilizing this paradigm in the investigation of incidence rates for invasive squamous-cell cancer in women who had at least one negative test, have been assembled

in Hakama et. al.[12] . The followers of this paradigm frequently speak of "the protective effect" of an interval of length x following y negative tests. The studies include cohort and case-control designs, and include data from Scotland, Iceland, Denmark, Norway, Sweden, Switzerland, Italy, and Canada. The present study will concentrate on screen-detected cases of CIS rather than incidence cases of invasive cancer since there are few cases of invasive cancer int the data available. However, analysis of CIS is of interest in its own right in that one of the postulated benefits of PAP test screening is the prevention of CIS with the attendant morbidity associated with its treatment.

As previously mentioned, the Pap test is scored on a scale from Class 1 to Class 4, with Class 1 representing no evidence of abnormality, i.e. a negative test result. However, there are discrepancies in the literature concerning the definition of "negative" tests. Boyes et al.[1] used a complicated algorithm to identify negative tests. The test result had to be either a Class 1 or a Class 2 which did not meet any of the following conditions: (a) one of three successive Class 2's; (b) one of a pair of Class 2's separated by an interval of ten months or more; or (c) followed directly by "a test of Class 3 or higher, or by a histological demonstration of dysplasia, carcinoma *in situ*, or invasive carcinoma". The majority of studies collected in Hakama[11] adapted the criteria used by Boyes et. al.[1] "A negative smear is either class I; class II followed by a class-I smear; or class II followed by a class-II smear within ten months, followed by a class-I smear"[11]. Some of the studies reported in Hakama adopted a liberal definition of "negative" in that "a negative smear is recorded when neither the cytological nor the

clinical examination leads to further cytological or gynaecological examination apart from succeeding screenings"[13].

Most cases of CIS are discovered as a result of follow-up tests to abnormal screen results. PAP tests taken in this context are called "diagnostic" as opposed to "screens". The frequency of such diagnostic tests is a different issue from screening frequency. According to Berrino et. al.[14]

> A usual way of coping with this problem has been to exclude diagnostic smears from the analysis (Clarke & Anderson[15]; La Vecchia *et al*[7]) or to exclude all the positive smears from both the series of cases and of controls (MacGregor *et al*[17]).

Berrino *et al.* include all tests taken "before the onset of symptoms". Boyes *et al* collapsed any such series with intervals less than 10 months into a single diagnostic test. Thus the results of different studies must be considered in light of the operational definitions of basic constructs such as "negative" tests and screening intervals, as the outcomes may hinge on these.

One of the case-control studies included in the Hakama collection, conducted by Vecchia et. al.[7], included a group of 145 women with cervical intra-epithelial neoplasia (CIN), (a new classification category that combines dysplasia and CIS stages), who were recruited from women referred to a university gynaecology clinic or the National Cancer Institute of Milan "for routine cervical screening". That is, they were detected by routine screening. The authors refer to a "diagnostic" test for the CIN cases. This presumably means an abnormal test result which leads to a positive diagnosis. Twenty-

three percent of the cases were classified "histologically" as "mild dysplasia" (CIN I), 26% as moderate dysplasia (CIN II), and 51% as severe dysplasia or carcinoma *in situ* (CIN III). The control group consisted of "women found to have normal cervical smears at the same screening clinics where CIN subjects had been identified. They were also matched for age by 5-year intervals". It is not clear what definition of "normal" was being used, although one of the tables suggests that class 2 tests were considered "normal". Since the authors were interested in the effectiveness of different patterns of routine screening, any tests obtained for the purposes of diagnosis "because of bleeding or other symptoms suggestive of cervical neoplasia" were excluded. "Subjects were specifically asked whether they had been screened by Pap tests, the number of times they had been screened, and their age at the first, last, and any abnormal test. The reference point for timing appears to be date of diagnosis for the cases and date of interview for the controls, (i.e., the exact definitions are not clear to me). "The odds ratios (and 95% confidence intervals) obtained for various risk factors comparing CIN cases and controls are presented in Table 2. (Evidently the index category is no previous screens).

The results of logistic regression analyses controlling for "age, social class, number of visits to doctor or clinic, number of sexual partners, age at first intercourse, education, cigarette smoking, number of previous hospital admission, oral contraceptive use, and history of cervical ectopia" are comparable to the univariate results presented in the table. Controlling for the total number of tests resulted in an odds ratio of 0.37 (95% CI 0.13-1.03) for last tests ≤ 5 years versus >5 years ensuring that the effect was not

Table 2. Odds ratios for risk factors comparing CIN cases with controls.

Risk Factor (relative to no previous smear)	Odds Ratio	95% CI
1 previous smear	0.27	0.10-0.71
$\geq$2 previous smears	0.12	0.06-0.25
<3 years since last smear*	0.09	0.04-0.20
$\geq$3, <5 years since last smear	0.31	0.11-0.85
$\geq$5 years since last smear	0.45	0.20-0.70
Risk Factor (vs class I/II smears only)	**Odds Ratio**	**95% CI**
No previous smear	11.76	5.59-24.75
One or more abnormal smears within one year of diagnosis/interview***	∞	-
One or more abnormal smears outside of a year prior to diagnosis/interview	7.18	2.83-18.19

* Excluding cases with a positive smear less than one year prior to diagnosis (not including "diagnostic smear") results in an odds ratio of 0.07

*** Relative to "normal smears only (class I/II)"

entirely due to the total number of tests. In summary,

> Screening on one occasion, irrespective of woman's age and time since the smear was taken, reduced the risk of ... CIN to about a quarter (RR=0.27). The degree of protection increased with increasing number of previous smears, and with decreasing interval since last smear, both trends in risk being highly significant. [And] women with previous positive smears remain at increased risk of cervical neoplasia.

The authors offer the following explanation for the findings with respect to CIN:

> The finding that screening reduces the risk of CIN may seem

> surprising, because strictly speaking Pap smears do not protect against the development of CIN. They are used to detect CIN, which if destroyed, may not develop into invasive cancer. The explanation for reduction in risk with increasing number of smears could be that women with a healthy cervix have more opportunity during their lifetime than do those with disease to accumulate multiple tests.

Berrino[16] offers an alternative explanation.

> Since CIN is a long-lasting disorder it is unlikely that women with newly diagnosed CIN would have been screened recently; if so, their CIN would have been detected then. Thus, previously screened women are bound to be under-represented among CIN cases detected in any given period. This bias may easily explain the observed association and its quite strong temporal trend without postulating any protective effect of screening.

Berrino seems to be referring to the effect of length of screening interval on prevalence under constant incidence. The groups defined by interval length are being compared on the odds of disease which reflects prevalence rather than incidence of disease. The odds ratios which La Vecchia et. al. report are "prevalence" odds ratios as opposed to incidence rate ratios which are the usual indices of relative risk. In some situations prevalence odds ratios approximate relative risk, but not when the factor under consideration is interval length. This subject will be elaborated below.

It should also be observed that PAP tests *could* serve to prevent the development of CIN, at least some forms of CIN. The diagnostic category "CIN" covers a variety of stages of disease, and although all of them are without symptoms and thus can only be screen detected, protection against advanced stages such as carcinoma *in situ* can be achieved by detection of earlier stages through screening and subsequent treatment.

Another case-control study reported in the Hakama collection was carried out by MacGregor et. al.[17] following the same paradigm of assessing the relative risk of disease as a function of time since last negative test and the number of previous negative tests. They recruited all cases of invasive squamous carcinoma of the cervix which appeared on the cancer registry in the Grampian region of Scotland between 1968 and 1982 who had attended for screening at least once. Eighty of the cases so obtained were identified as having been "detected by routine screening". It should be emphasized that the cases were invasive cases and thus probably quite distinct from CIS cases with respect to properties like duration of disease. The following method was used to sample controls:

> Five random controls, matched for year of birth and with the additional constraint that each must have entered the study (at the time of her first negative smear) before the date of diagnosis of cancer in the patient [and have] been screened within six months either side of the date of the screening test at which the cancer was diagnosed. This was to ensure that both patients and controls were drawn from the same population of women - namely, those attending for routine (asymptomatic) screening.[17]

The results are reported in terms of "relative protection", "the inverses of the relative risks". "The relative protection decreased progressively with increasing time since last negative smear"[17]. This study would seem to suffer from the same problem of interpretation as the preceding one, at least for the screen-detected cases, namely, since the "exposure" variable, "time since last negative test", is related to the duration of disease, the prevalence odds ratio is not equal to the incidence odds ratio[18].

One of the requirements of a valid case-control study is that selection of controls be

independent of the exposure variable[18] status. The MacGregor et. al. study uses different criteria to select cases and controls. The cases are required to have had one "diagnostic" screen within a 15 year period and at least one screen prior to that, whereas the controls for a given case are required to have had one screen prior to and another screen within a year of the case's "diagnostic" screen. The issue is whether these procedures for selecting cases and controls have the same, if any, sampling biases with respect to screening frequency, which is one of the "exposure" variables under consideration. We will subsequently examine the implications of these methods of selecting cases and controls under three theoretical models.

Not all studies have reported the same inverse relationship between risk of disease (invasive or otherwise) and screening frequency. For example, van Oortmarssen & Habbema[19] explain the high incidence rate in the first two years following a negative test for the B.C. data as possibly being due to the testing of symptomatic individuals because the "screening program" started as a diagnostic support service. The Manitoba study, also reported in Hakama[20], proposed different explanations for different age groups for "the lack of a trend towards increasing risk of developing invasive disease with increasing time since a negative smear".

> The lack of a trend towards increasing risk of developing invasive disease with increasing time since a negative smear may be attributable to different reasons depending on the age during which the woman-years occurred. In the younger women (<35 years of age), much of the long-term follow-up, especially after only one negative smear, is in error because many women may have changed their names, and subsequent smears were recorded under new names. This would account for the low incidence after only

one smear among women under 35 years of age.

> Among women under 40 years of age (after a five-year interval, 45 years of age) in particular, there is a relatively high rate of migration, and the calculation of woman-years at risk does not take this into account. Thus, the woman-years will be overestimated, and the incidence rates underestimated, by progressively greater amounts with increasing time interval; and there will be a significant effect after longer time intervals, when the woman-years at risk are already small.
>
> A high rate of hysterectomies was experienced in Manitoba among women aged between 40 and 50 years of age, in particular during the years 1969-1975. Hysterectomy would have the same effect as migration of underestimating incidence in a progressive manner with time. Among screened women over 55 years of age, a large proportion of cases may be due to false negatives, since the incidence of in-situ cervical cancer in women at these ages is low. A smaller proportion of women of these ages has been screened, and the low sensitivity of the test for detecting invasive cases among women who were screened because of symptoms would have a considerable effect in masking a trend.[20]

Some screening programs are more organized than others in the sense that there is an effort to screen the entire population at regular intervals. Others tend to rely more on individual choice and thus are more prone to selection biases. Some of the Scandinavian countries appear to have implemented relatively more organized screening programs where an attempt is made to screen all individuals at risk at regular intervals. The results from the "organized" screening programs tend to indicate increasing incidence rates of invasive cancer with increasing intervals and the increase is more gradual with more previous negative tests. The "opportunistic" screening programs like B.C.'s or Manitoba's do not show the same pattern. In fact the B.C. data suggests higher incidence rates of CIS are associated with shorter screening intervals[21], which is

contrary to the hypothesis that frequent screening reduces the risk of developing carcinoma *in situ*.

The cohort studies reported in Hakama et al. estimated incidence rates of invasive cancer by age, number of previous negative tests, and time since last negative tests by dividing the number of cases observed in each category by the total number of woman-years at risk observed for each category[22]. Apparently no attempts have been made to fit multivariate models to incidence data. The present study will fit a Poisson regression model, also called a log-linear model, which is a particular case of generalized linear models[30] The Poisson model is appropriate for modelling counts of independent events under a Poisson-like process with a constant rate. The canonical model, to be use here, assumes that the covariate effects are multiplicative with respect to the expected number of counts. The logarithm of counts observed in the covariate classes are fit to a linear function of the covariates with an optional "offset" which is given a coefficient of one. If, for this example, the offset is taken to be the log of the time-at-risk accumulated in a covariate class, then the log-linear model effectively models the log of the estimated incidence rate as a linear function of the covariates. If Y_i represents the number of incidence cases in a covariate class defined by a covariate vector $\mathbf{x}_i$ then the log-linear model can be written in the form[30]

$$\log\{E(Y_i)\} = \text{offset} + \beta' x_i.$$

The method of analysis used in the case-control studies consisted primarily of logistic regression with conditional likelihood functions[11]. Logistic regression is also a special case of generalized linear models. If π_i represents the probability of disease for the i^{th} covariate class, then $\pi_i/(1-\pi_i)$ is the odds of disease. Logistic regression models the log of the odds of disease as a linear funcion of the covariates. Using the same notation for covariates as in the case of Poisson regression, this is expressed symbolically as

$$\log(\frac{\pi_i}{1-\pi_i}) = \beta' x_i.$$

The adaptation of the logistic model to case-control studies involves conditioning on an individual being sampled for the study. Breslow and Day[26] demonstrate using Bayes' theorem, that if sampling is independent of exposure status then the model coefficients for the exposure variables are the same as for the population as a whole. When controls are matched with a particular case, the method of conditional likelihoods can be used to eliminate nuisance parameters. This general principle of statistical inference is discussed, for example, in Cox and Hinkley, 1974[23]. The unconditioned likelihood function may involve parameters, such as age, which are known to be important covariates, but which are not of interest in the given study, i.e., they are nuisance parameters. The method of conditional likelihood replaces the unconditional likelihood with a conditional likelihood which conditions on the nuisance parameters. Estimates of the parameters of interest are obtained from the maximum likelihood solution of the conditional likelihood equation. With one-to-many case-control matched designs, a likelihood function for each matched set is constructed which conditions on the fact that exactly one of the members

of the set is a case. Following the presentation in Armitage[24], if the probability of disease of an individual in a matched set, s, is given by $\exp(\alpha_s+\beta x_i)$, $i=0,1,\ldots,c$ where c is the number of controls in each matched set, then the probability that the observed case is diseased given that exactly one member of the set is diseased is given by,

$$\frac{\exp(\alpha_s+\beta x_0)}{\sum \exp(\alpha_s+\beta x_i)},$$

$i=0,1,\ldots,c$. The term $\exp(\alpha_s)$ factors out of the numerator and denominator giving,

$$\frac{\exp(\beta x_0)}{\sum \exp(\beta x_i)}.$$

The conditional likelihood is then the product of such terms from each matched set.

Chapter 3

Epidemiological Methods

The following is a review of the basic principles of sampling in case-control studies[18]. One first selects a case series. Often these comprise all cases reported in a given period. It is recommended, however, that the cases form an "etiologically" homogeneous group, i.e. they probably have a common causal history of disease development. Controls are selected from a pool of eligible controls. The crucial criterion for eligibility is that the individual would have been included as a case had they developed the disease. Controls can be matched with cases or not. Rothman[18] demonstrates how the odds-ratio obtained from a case-control study provides an estimate of relative risk as follows:

> The relevant data on disease incidence for a time period of length t might be summarized as

$$I_1 = \frac{a}{P_1 t},$$

and

$$I_0 = \frac{b}{P_0 t},$$

where I_1 and I_0 are the incidence rates among exposed and unexposed, respectively, a and b are the respective numbers of individuals who developed disease during time interval t, and P_1 and P_0 are the respective population sizes. ... The cases in a case-control study are the individuals who became ill during the time period, that is a total of $(a + b)$ individuals. ... If a proportion, k, of the combined exposed and unexposed cohorts is taken as controls, and the number of such controls is c for exposed and d for unexposed, then the incidence rates among exposed and unexposed could be estimated as

$$I_1 = k\frac{a}{ct},$$

[Actually this is an approximation which assumes that the risk of disease is small. The correct estimate is $a/\{t(c/k+a)\}$.]

and [continuing with Rothman's presentation]

$$I_0 = k\frac{b}{dt},$$

... the relative incidence, or rate ratio (RR, often referred to as *relative risk*), is obtained as

$$RR = \frac{I_1}{I_0} = \frac{ad}{bc}.$$

Since the sampling fraction, k, is identical for both exposed and unexposed, it divides out, as does t. The resulting quantity, ad/bc, is the exposure odds ratio (ratio of exposure odds among cases to exposure odds among controls), often referred to simply as the *odds ratio*. This cancelation of the sampling fraction for controls in the odds ratio thus provides an unbiased estimate of the incidence rate ratio from case-control data [Sheehe[25]; Miettinen[26]]. The central condition for conducting valid case-control studies is that controls be selected independently of

> exposure status to guarantee that the sampling fraction can be removed from the odds ratio calculation.[18]

Breslow and Day[26] give the same caution.

> One fundamental sampling requirement to which attention is drawn is that the *sampling fractions for cases and controls must be the same regardless of exposure category*. If exposed subjects are more or less likely to be included in the sample than are the unexposed, serious bias can result.

The ultimate aim of many epidemiological investigations is to estimate the relative risk of contracting a disease in a given time period for those exposed to some condition in comparison to those not exposed. Relative risk expresses the ratio of risk rates for two groups, i.e., relative risk is the ratio of the probabilities of developing disease for the exposed and unexposed groups. The odds ratio is the ratio of odds of contracting disease for those exposed relative to those not exposed, where the odds of contracting disease is defined as the probability of contracting disease divided by the probability of not contracting disease. For rare diseases, the probability of not contracting disease is close to unity, so the odds ratio reduces to the relative risk.

The probability of contracting disease in a given time period, $P(t)$, can be approximated by the cumulative incidence rate. This is demonstrated by Breslow and Day[27] who credit Elandt-Johnson[28] with the following expression for the instantaneous incidence rate, $\lambda(t)$,

$$\lambda(t) = \frac{1}{1-P(t)} \times \frac{dP(t)}{dt},$$

and hence,

$$1-P(t) = \exp\{-\Lambda(t)\}, \tag{3.1}$$

where,

$$\Lambda(t) = \int_0^t \lambda(u)du,$$

the cumulative incidence rate. Taking logarithms gives,

$$\Lambda(t) = -\log\{1-P(t)\} \approx P(t),$$

when the disease is rare or t is small. Relative risk is defined as the ratio of incidence rates in exposed versus non-exposed individuals. Following the presentation in Breslow and Day, if $r = \lambda_1/\lambda_2$, the ratio of two incidence rates, which is the definition of *relative risk* adopted by Breslow and Day, is constant over a period of time t, then from (3.1) we have,

$$
\begin{aligned}
P_1(t) &= 1 - \exp\{-\Lambda_1(t)\} \\
&= 1 - \exp\{-\int_0^t \lambda_1(u)du\} \\
&= 1 - \exp\{-r\int_0^t \lambda_2(u)du\} \\
&= 1 - [\exp\{-\Lambda_2(t)\}]^r \\
&= 1 - \{1-P_2(t)\}^r \\
&\approx rP_2(t)
\end{aligned}
$$

providing the disease is rare or the time period is short. Breslow and Day observe that "In general, the ratio of disease risks is slightly less extreme, i.e., closer to unity, than is the ratio of the corresponding rates"[26].

The usual way of estimating incidence rates in a cohort study is to divide the number of cases occurring in a given time period by the number of "person-years" at-risk contributed by the cohort population. Person-years is the total of all time at-risk from all individuals. To estimate incidence rates for strata which are time dependent, i.e., an individual might belong to one strata level at one point in time and to another strata level at another time, the usual procedure is to assign the case to the strata level at which the disease "occurs", and to partition the individual's total time at risk among the strata according to time spent in the strata. The principle at work is stated by Breslow and Day as follows,

> The correct assignment of each increment in person-years of follow-up is to that same exposure category to which a death

would be assigned should it occur at that time.

This procedure is difficult to apply for screen-detected diseases since the exact time of disease incidence is unknown. All that is known is that the disease occurred at some time between the start and the end of the screen interval. If the disease is rare, this fact will not have much impact on the denominator estimates since they are comprised mostly of intervals which do not result in diagnosis. The problem is in how to classify the cases, i.e., which numerators to increment, with respect to time dependent covariates, e.g. age, date, etc.

One method which has been used is to date the onset at the midpoint between the date of diagnosis and last negative screen prior to that. The time at risk is partitioned over time dependent categories as if the case were incident at the midpoint. However this procedure violates the principle that time at-risk should be applied to that category to which the case would have been assigned had it occurred at that time. If the case is assigned to the category containing the midpoint and if portions of time-at-risk are assigned to categories according to overlap, then portions of the interval which overlap categories other than the one containing the midpoint will be assigned to categories other than the one to which the case would have been assigned had disease occurred at that time. Whatever method one chooses to locate the time of disease (e.g. interval midpoint), the principle would seem to imply that the at-risk portion of the interval should be applied to the person-years of the same category as the case. How is the

principle to be applied to control intervals? The time at risk should be applied to the same category that the case would have been assigned had disease occurred. This would seem to imply that the entire interval should be assigned to one category, namely the one to which the case would have been assigned had disease occurred during the interval.

This is perhaps the alternative method employed in some analyses by Boyes et. al., namely,

> The denominators are obtained by calculating the number of years at risk between pairs of successive smears and allocating this number to the age groups within which the mid-point occurs. These years of risk are then aggregated over all pairs of smears and all women to produce appropriate denominators, expressed as person-years of risk which can be used in the calculation of the event rates.[1]

This description is ambiguous in that reference is made to "the age groups", i.e. more than one age group, within which the mid-point occurs. But assuming that the entire interval is assigned to the unique age group containing the midpoint, this will mean contributions being made to one category from times when individuals actually belong to another category. The categories will tend to be blended. But at least the method does not violate the principle of assigning time-at-risk to the category to which the case would have been assigned had disease occurred.

If instead of counting the number of incident cases in a given time period one were to count the number of individuals with disease at a given point in time and divide this by the number without disease, we would be estimating the prevalence odds of disease[18].

The prevalence odds is a function of the average duration of disease, since the number of individuals found with disease at a given point in time is equal to the number who contracted the disease in the past and still have it. Under certain circumstances the prevalence odds approximately equals the incident rate multiplied by the average duration of disease[18]. Rothman gives the following presentation. First the incidence rate equals the inverse of the average time until incidence. To see this, imagine following a finite population until everyone gets disease. At that point the incidence can be calculated by the number of people divided by the sum of their waiting times. Invert this and you get the average waiting time. Assume the population is in a steady state, the number of people entering the disease pool equals the number of people leaving it, and let N be the total number of people, P the number with disease, I the incidence rate, I' the incidence rate of exiting from the disease pool, then for any time interval Δt we have, by the steady state assumption that,

$$I\Delta t(N-P) = I'\Delta tP.$$

And hence, since the mean duration in a state equals the reciprocal of the incidence rate for exiting from that state, i.e., either going from a state of health to a state of disease or vice versa, if $\bar{D}$ represents the average duration of disease, then $\bar{D}=I'^{-1}$, and hence,

$$I\Delta t(N-P) = (1/\bar{D})\Delta tP \Rightarrow \frac{P}{N-P} = I\bar{D}. \tag{3.2}$$

That is, the prevalence odds equals the product of the incidence rate and the average duration of disease.

If the prevalence rate is small then $N\text{-}P \approx N$ and the prevalence odds is approximately equal to the prevalence rate which in turn will be equal to the incidence rate (for a steady state population) when the average duration of disease is one unit of time. Prevalence odds ratios equal incidence rate ratios i.e. relative risk, when the average duration of disease is equal in the two exposure groups under comparison. Or, alternatively, if the average duration of disease is estimable for the groups then the prevalence odds ratio could be adjusted accordingly to provide an estimate for the incidence ratios, i.e. relative risk.

That case-control studies of screen-detected diseases involve prevalence odds ratios rather than incidence odds ratios has been reported by Sasco et. al.[29], but the implications in terms of the findings of such studies as MacGregor et. al. and La Vecchia et. al. have not been articulated except, perhaps, in Berrino's interpretation of La Vecchia's findings.

Chapter 4

Screening Models

This study addresses two issues raised by case-control studies involving screen-detected disease, namely the effects of proposed methods of sampling controls and the use of prevalence odds ratios to estimate relative risk. We examine these issues in the context of three screening models: a fixed interval model which is oversimplified but which serves to illustrate some of the issues involved; a uniformity model which is less restrictive than the fixed interval model; and a Poisson model. The primary objective is to see whether the proposed method of matching controls, based on having a screen within six months of the "date of diagnosis" of a case, affects the resulting odds ratio. Since the independent variable, interval length, may be related to the probability of being sampled as a control, there is a possibility of sampling bias. Interval length is also related to duration of disease, so by the theory developed around equation (3.2), the odds ratio, which is a prevalence odds ratio, may not approximate relative risk.

We will assume that the cases consist of all cases of disease diagnosed within a study period of length L. The controls will be assumed to be matched with the cases on the

basis of having a screen within a matching period of length l centred around the date of the screen which led to diagnosis for the matched case. In the study by MacGregor et. al.[17] L was 15 years and l was one year. The requirements for controls must be modified at the two ends of the study period. Since a prerequisite for eligibility as a case or a control is that the screening interval ends within the study period. Controls must also be at-risk for disease, i.e. currently free of disease and without prior treatment that would preclude disease. To simplify matters we will assume that the tests are error-free and that the disease under consideration does not regress. The null hypothesis is that screening does not have any effect other than identifying the presence of disease.

Fixed Interval Model

Consider a population which consists of two groups in equal numbers, say 100,000 each, one group of individuals who are screened annually and another who are screened bi-annually (i.e., every two years), with 50,000 screened in each year. Assume also that the incidence rates are constant, small, and equal in the two groups, say 1 per 1000 per year, and that any diagnosed individuals are immediately replaced by others from an external source. Suppose each case occurring within a 2 year period is selected and matched with a control drawn randomly from those tested in the same year the case is diagnosed. Assume diagnosis only occurs as a result of a screen and is immediate. The situation would be as depicted in Table 3.

The reason 100 diagnosed cases occur in the bi-annual screening groups on each testing occasion is that 50 were incident during the past year while another fifty were left over

Table 3. Hypothetical example of sampling matched controls under fixed interval screening.

screening pattern	odd year				even year			
	screened	incidence cases	diagnosed cases	sampled controls	screened	incidence cases	diagnosed cases	sampled controls
annually	100,000	100	100	133.5	100,000	100	100	133.5
even years	0	50	0	0	50,000	50	100	66.5
odd years	50,000	50	100	66.5	0	50	0	0

from the previous year since they were not screened and therefore not diagnosed at the time. Two hundred controls are sampled randomly from the 150,000 who were screened in a given year, resulting in a 2-1 ratio of annual to bi-annual screeners among the controls. If we now compute the odds ratio for screening frequency we have the setup shown in table 4.

Table 4. Odds ratio for annual vs bi-annual screeners.

	annual	bi-annual
cases	200	200
controls	267	133
Odds ratio $= 134/267 \approx 1/2$		

The incidence risk ratio is unity, but the case-control methodology employed results in

an odds ratio of 0.5. The inference we would draw from the results of this study, under normal circumstances, would be that annual screening reduces the risk of disease.

Uniformity model

The preceding example assumed fixed screening intervals. This is probably not realistic since interval lengths vary within individuals. In the uniformity model we do not assume any connections between intervals within an individual, however, we do not rule them out either. We only assume that the distribution of screening intervals is stationary, i.e., the same at any point in time, and "uniform", i.e., the starting times for intervals of a given length are uniformly distributed over a period of time which encompasses all possible starting times of intervals that end in the study period. Thus in the uniformity model, the units of analysis are intervals with lengths and starting times which satisfy the stationarity and uniformity conditions. This model, like the fixed interval model, is not intended to be realistic but, rather, hypothetical for the purposes of illustration. We do not attempt to develop a realistic model of screening patterns.

Assuming a constant incidence rate of disease I, the probability that an individual develops disease during an interval of length $U=u$ is approximately Iu, for if D is a random variable which takes the value 1 when disease occurs during an interval and 0 otherwise, then,

$$P(D=1 \mid U=u) = 1-\exp\{-\Lambda(u)\} \approx Iu. \tag{4.1}$$

Assume that U is a continuous random variable representing interval length with density

f_U, then by Bayes' theorem and (4.1), the conditional density of U given $D=1$, $f_{U \mid D=1}$, is given by,

$$f_{U|D=1}(u) = \frac{P(D=1 \mid U=u) \cdot f_U(u)}{\int_{u=0}^{\infty} P(D=1 \mid U=u) \cdot f_U(u) du}, \tag{4.2}$$

$$\approx \frac{Iu \cdot f_U(u)}{\int_{u=0}^{\infty} Iu \cdot f_U(u) du},$$

$$= \frac{u \cdot f_U(u)}{E(U)}.$$

This gives the distribution of interval lengths for cases. Hence, the odds for intervals of length u conditional on intervals having length u or v and disease occurring is given by,

$$\frac{f_{U|D=1}(u)}{f_{U|D=1}(v)} \approx \frac{u \cdot f_U(u)}{v \cdot f_U(v)}. \tag{4.3}$$

Next we consider screening histories containing intervals that end within a matching period of length l centred at the time of diagnosis of a case. The matching period is centred at the screening date which "led" to the diagnosis of a case. The case-control methodology which is under examination requires cases to be matched with controls who have a screen within specified limits of the "diagnostic" screen. Since an individual may have more than one interval ending in the matching period, for mathematical convenience and to ensure uniqueness we will require control intervals to span the start of the matching period, i.e. begin before and end after the start of the matching period. The event of an interval spanning a point in time t will be represented by a random variable

S which takes the value 1 if an interval spans t, and 0 otherwise.

Let the matching period for the selection of controls for a particular case begin at time t and have length l. The probability that the endpoint of an interval which spans t lands in the matching period follows from the uniformity assumption. Since the starting point of an spanning interval of length u is uniformly distributed over $(t\text{-}u,\ t)$ and the sub-interval of starting points which land the endpoint in $(t,\ t+l)$ has length l while the entire interval of possible starting points has length u, the probability of the endpoint landing in $(t,\ t+l)$ is $\min(1,\ l/u)$. The event of ending in the matching period is represented by the random variable M which takes the value 1 when an interval ends in the matching period and 0 otherwise. Then we have,

$$P(M=1 \mid U=u,\ S=1) = \min(1,\ l/u). \tag{4.4}$$

The probability of an interval having length u given that it spans t ($S=1$), ends in the matching period ($M=1$), and doesn't lead to diagnosis ($D=0$) is given by,

$$\begin{aligned} f_{U|M=1,S=1,D=0}(u) &= \frac{P(M=1,\ S=1 \mid U=u,\ D=0) \cdot f_{U|D=0}(u)}{P(M=1,\ S=1 \mid D=0)}, \\ &= \frac{P(M=1,\ S=1 \mid U=u) \cdot f_{U|D=0}(u)}{P(M=1,\ S=1 \mid D=0)}, \end{aligned} \tag{4.5}$$

by Bayes' theorem and the independence of interval span and endpoint location from disease status. The conditional probability of spanning t and ending in the matching period, given that the interval has length u is

$$P(M=1,\ S=1 \mid U=u) = P(M=1 \mid S=1,\ U=u) \cdot P(S=1 \mid U=u). \tag{4.6}$$

We already have an expression for the first factor, namely $\min(1, l/u)$. For the probability that an interval of length u spans a time t we can assume that $t \in (0, L-l)$ if cases diagnosed within $l/2$ of the end of the study period are matched with controls with intervals ending within l of the end of the study period. Also note that the interval endpoint can range from 0 to L, but not beyond for otherwise if disease were detected at an endpoint beyond L the individual would not qualify as a case in the study. By the assumption of uniform starting times for intervals of length u over any specified period, it follows that the endpoints of intervals of length u are uniformly distributed over $(0, L)$. Therefore, the probability of spanning a point t in $(0, L-l)$ is given by the ratio of the length of the region containing favourable endpoints to the length of all endpoints (L). The length of favourable endpoints depends on the relative sizes of u and l, for if $u>l$ then for $t \in (L-u, L-l)$ the range of endpoints for which the interval spans t will be limited to $(L-t, L)$. Otherwise the range of favourable endpoints will have length u. These results are summarized as follows,

$$P(S=1 \mid U=u) = \begin{cases} \dfrac{u}{L}, & 0 \le t < L-u,\ u>l \\ \dfrac{L-t}{L}, & L-u \le t \le L-l,\ u>l \\ \dfrac{u}{L}, & l \ge u \end{cases} . \tag{4.7}$$

Substituting $\min(1, l/u)$ for $P(M=1 \mid S=1, U=u)$ and (4.7) in (4.6) gives,

$$P(M=1,\ S=1 \mid U=u) = \begin{cases} \frac{l}{u}\cdot\frac{u}{L} = \frac{l}{L}, & 0\le t<L-u,\ l<u \\ \frac{l}{u}\cdot\frac{L-t}{L}, & L-u\le t<L-l,\ l<u \\ \frac{u}{L}, & l\ge u \end{cases} . \tag{4.8}$$

Next we compute

$$\begin{aligned} f_{U|D=0}(u) &= \frac{P(D=0 \mid U=u)\cdot f_U(u)}{\int_0^\infty P(D=0 \mid U=u)\cdot f_U(u)du}, \\ &\approx \frac{f_U(u)}{\int_0^\infty P(D=0 \mid U=u)\cdot f_U(u)du}, \\ &\approx f_U(u), \end{aligned} \tag{4.9}$$

since $P(D=0 \mid U=u) \approx 1\text{-}Iu \approx 1$. This assumes that if $[a,b]$ is the support of f_U then $P(D=0 \mid U=u) \approx 1\text{-}Iu$ is true for all $u \in [a,b]$. Substituting (4.8) and (4.9) in (4.5) gives,

$$f_{U|M=1,S=1,D=0}(u) \propto f_U(u)\times \begin{cases} \frac{l}{L}, & 0\le t<L-u,\ l<u \\ \frac{l(L-t)}{uL}, & L-u\le t<L-l,\ l<u \\ \frac{u}{L}, & l\ge u \end{cases},$$

which shows that the distribution of intervals which span t and land in the matching period approaches the distribution of intervals starting from a point t as $l \downarrow 0$. Now dividing by the same expression for v in place of u, and assuming without loss of generality that $u < v$ gives,

$$\frac{f_{U|M=1,S=1,D=0}(u)}{f_{U|M=1,S=1,D=0}(v)} \approx \frac{f_U(u)}{f_U(v)} \times \begin{cases} 1 \ , & 0 \le t < L-v, \ \ l < u \\ \frac{u}{l} \ , & 0 \le t < L-v, \ \ u \le l < v \\ \frac{v}{L-t} \ , & L-v \le t < L-u, \ \ l < u \\ \frac{uv}{l(L-t)} \ , & L-v \le t < L-l, \ \ u \le l < v \\ \frac{v}{u} \ , & L-u \le t < L-l, \ \ l < u \\ \frac{u}{v} \ , & l \ge v \end{cases} \tag{4.10}$$

This expression gives the odds for intervals of length u which meet the selection criteria for controls when attention is restricted to intervals of length u or v assuming the regularity conditions hold. Hence under these assumptions the ratio of the odds for cases (4.3) to odds for matched controls (4.10) gives the (prevalence) odds ratio, Ψ,

$$\Psi = \frac{u}{v} \times \begin{cases} 1 \ , & 0 \le t < L-v, \ \ l < u \\ \dfrac{u}{l} \ , & 0 \le t < L-v, \ \ u \le l < v \\ \dfrac{v}{L-t} \ , & L-v \le t < L-u, \ \ l < u \\ \dfrac{uv}{l(L-t)} \ , & L-v \le t < L-l, \ \ u \le l < v \\ \dfrac{v}{u} \ , & L-u \le t < L-l, \ \ l < u \\ \dfrac{u}{v} \ , & l \ge v \end{cases}.$$

Thus under the given regularity assumptions, the prevalence odds ratio will equal the ratio of interval lengths ignoring complications that arise within an interval length of the end of the study period, which may be a sizable period depending on the intervals under comparison. It might be advisable to stratify the analyses in comparing different interval lengths by the locations of the case diagnosis dates.

Ignoring complications arising near the end of the study period, the distribution of interval lengths given that they span a point in time t is given by,

$$f_{U|S=1}(u) = \frac{P(S=1 \mid U=u) \cdot f_U(u)}{\int_0^\infty P(S=1 \mid U=u) \cdot f_U(u) du},$$

$$\approx \frac{u \, f_U(u)}{E(U)},$$

since $P(S=1 \mid U=u) \approx u/L$. This is the same as the distribution derived earlier for case intervals (equation 4.2). But the controls are required to satisfy the additional constraint of landing in an interval of length l which we have seen effectively weights the distribution of controls inversely by interval length.

Poisson Model

PAP tests may be considered rare events, and if screening intervals are independent of past intervals, then the screening process may be modelled by a Poisson process. We will assume the Poisson process is time homogenous with parameter μ. This implies that interval lengths have an exponential distribution with mean $1/\mu$. Once the process has been going for a while the distribution of starting times of intervals relative to the beginning of the process will be distributed as the sum of independent exponentials. Hence the distribution of starting times within a given small interval will be approximately uniform. Thus the Poisson model approximately satisfies the assumptions of the uniformity model in the previous section. However, the Poisson model specifies the distribution of the interval lengths. Once again, the objective of this excercise is to derive the odds ratio which results from selecting controls to match cases on the basis of having a screen within a specified time of the date of diagnosis under the null hypothesis.

I. Distribution of Spanning Intervals

We will first calculate some useful distributions which will be needed for the calculation of case and control screen interval distributions. As before, let t be the start of a matching period. We now define two random variables. Let T_0 be the time of the last test before the start of the matching period, and let T_1 be the time of the first test after the start of the matching period. Let $N(t)$ be the number of tests during time t. The

probability that there are n tests before t is given by,

$$P(N(t)=n) = \frac{e^{-\mu t}(\mu t)^n}{n!}. \tag{4.11}$$

For $n \geq 1$, $P(T_0\epsilon[t_0,t_0+dt_0],\ T_1\epsilon[t_1,t_1+dt_1] \mid N(t)=n)$

$$= \frac{\dfrac{e^{-\mu t_0}(\mu t_0)^{n-1}}{(n-1)!} \cdot \mu dt_0 \cdot e^{-\mu(t_1-t_0)} \cdot \mu dt_1}{\dfrac{e^{-\mu t}(\mu t)^n}{n!}},$$

$$= n\mu \frac{t_0^{(n-1)}}{t^n}\ e^{-\mu(t_1-t)}\ dt_0 dt_1.$$

The four factors in the numerator represent the probability of n-1 tests before t_0, one test in dt_0, no tests between t_0 and t_1, and finally, one test in dt_1. We will not consider $n=0$ since there is then no screen prior to t. Thus multiplying by $P(N(t)=n)$, for $n\geq 1$, and summing over n gives,

$$P(T_0\in[t_0,\ t_0+dt_0],\ T_1\in[t_1,\ t_1+dt_1],\ N(t)\geq 1) \tag{4.12}$$

$$= \sum_{n=1}^{\infty} n\mu \frac{t_0^{(n-1)}}{t^n} e^{-\mu(t_1-t)} \cdot \frac{(\mu t)^n e^{-\mu t}}{n!} dt_0 dt_1,$$

$$= \sum_{n=1}^{\infty} \frac{(\mu t_0)^{(n-1)}}{(n-1)!} \mu^2 e^{-\mu t_1} dt_0 dt_1,$$

$$= \mu^2 e^{-\mu(t_1-t_0)} dt_0 dt_1.$$

Let $U=T_1-T_0$, the length of the smear interval spanning t, then the Jacobian associated with the parameter transformation $(t_1,t_0) \rightarrow (u,t_0)$ is

$$J(t_0,t_1 \mid t_0,u) = \frac{\partial t_0}{\partial t_0}\cdot\frac{\partial t_1}{\partial u} - \frac{\partial t_1}{\partial u}\cdot\frac{\partial t_1}{\partial t_0} = 1\cdot 1 - 0\cdot 1 = 1 .$$

The marginal distribution for U is obtained by integrating (4.12) over t_0 where t_0 ranges from $\max\{(t\text{-}u),0\}$ to t, giving,

$$f_{U,N(t)\geq 1}(u) = \begin{cases} \int_0^t \mu^2 e^{-\mu u}\, dt_0, & u \geq t \\ \int_{t-u}^t \mu^2 e^{-\mu u}\, dt_0 , & u < t \end{cases} , \tag{4.13}$$

$$= \begin{cases} \mu^2 t e^{-\mu u} , & u \geq t \\ \mu^2 u e^{-\mu u} , & u < t \end{cases} .$$

Now as $t \to \infty$ (i.e., the interval spans a point distant from the start of the process) and $P(N(t)\geq 1)\to 1$, then,

$$f_{U,N(t)\geq 1}(u) \to f_U(u) = \mu^2 u e^{-\mu u}. \tag{4.14}$$

The interval length thus has a Gamma distribution (equal to the sum of two independent exponentials with parameter μ).

II. Distribution of Spanning Intervals Ending in the Matching Period.

We will now calculate the same distribution, i.e., the distribution of the interval from the last smear (T_0) prior to t to the time of the next smear, T_1, subject to the condition that $T_1 \leq t+l$, i.e., the endpoint of the spanning interval lands in the matching period. For $N(t) \geq 1$,

$$P(T_0 \epsilon [t_0, t_0+dt_0],\ T_1 \epsilon [t_1, t_1+dt_1] \mid N(t)=n,\ T_1 < t+l)$$

$$= \frac{\frac{e^{-\mu t_0}(\mu t_0)^{n-1}}{(n-1)!} \cdot \mu dt_0 \cdot e^{-\mu(t_1-t_0)} \cdot \mu dt_1}{\frac{e^{-\mu t}(\mu t)^n}{n!} \cdot (1-e^{-\mu l})},$$

$$= n\mu \frac{t_0^{(n-1)}}{t^n} \frac{e^{-\mu(t_1-t)}}{1-e^{-\mu l}} dt_0 dt_1.$$

Multiplying the terms by $P(N(t)=n)$ and summing gives,

$$P(T_0 \epsilon [t_0, t_0+dt_0],\ T_1 \epsilon [t_1, t_1+dt_1],\ N(t) \geq 1 \mid T_1 < t+l)$$

$$= \frac{\mu^2 e^{-\mu(t_1-t_0)}}{1-e^{-\mu l}} dt_0 dt_1 \qquad (4.15)$$

Again, let $U=T_1-T_0$, and integrate over T_0 to get the marginal distribution of U. The range of integration is determined by the possible values of t_0 under the constraint on t_0 and t_1 with respect to t. The integral may appear in two different forms depending on whether $l<t$ or $l>t$. We are typically only interested in the case $t>l$, i.e., where the length of the interval for having a smear is small compared to the time since the process

started. Thus, we have

$$f_{U,N(t)\geq 1|T_1\leq t+l}(u) = \begin{cases} \int_{t-u}^{t} \frac{\mu^2 e^{-\mu u}}{1-e^{-\mu l}}\, dt_0 \quad , & 0\leq u<l \\ \int_{t-u}^{t+l-u} \frac{\mu^2 e^{-\mu u}}{1-e^{-\mu l}}\, dt_0 \quad , & l\leq u<t \\ \int_{0}^{t+l-u} \frac{\mu^2 e^{-\mu u}}{1-e^{-\mu l}}\, dt_0 \quad , & t\leq u<t+l \\ 0 \quad , & u\geq t+l \end{cases} \tag{4.16}$$

Hence,

$$f_{U,N(t)\geq 1|T_1\leq t+l}(u) = \begin{cases} \frac{\mu^2 u e^{-\mu u}}{1-e^{-\mu l}} \quad , & 0\leq u<l \\ \frac{\mu^2 l e^{-\mu u}}{1-e^{-\mu l}} \quad , & l<+u<t \\ \frac{\mu^2 (t+l-u) e^{-\mu u}}{1-e^{-\mu l}} \quad , & t\leq u<t+l \\ 0 \quad , & u\geq t+l \end{cases} \tag{4.17}$$

We may now use this expression to derive the distribution of the screen interval length for spanning intervals required to fall within a matching interval. Let $t \to \infty$, then $P(N(t)\geq 1) \to 1$ and,

$$f_{U,N(t)\geq 1|T_1\leq t+l}(u) \to f_{U|T_1\leq t+l}(u) = \begin{cases} \frac{\mu^2 u e^{-\mu u}}{1-e^{-\mu l}} \quad , & u<l \\ \frac{\mu^2 l e^{-\mu u}}{1-e^{-\mu l}} \quad , & u\geq l \end{cases} \tag{4.18}$$

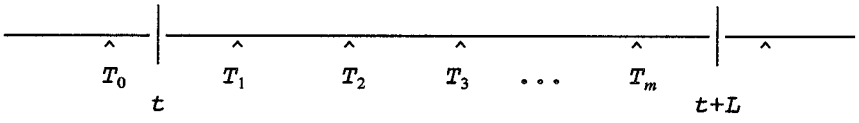
Note that if l is small in (4.18) above (i.e., $l \ll \mu^{-1}$) then for practical purposes we have

$$f_{U|T_1 \le t+l}(u) = \mu e^{-\mu u}, \tag{4.19}$$

and the distribution of intervals which terminate at t is the same as the overall distribution of intervals. This is the same result that was found for the uniformity model, and suggests that the method of sampling controls based on the criterion of having a screen close to the date of diagnosis of a case, does not affect the distribution of interval lengths.

III. Risk of Disease Over the Study Period.

The preceding distributions are not yet directly applicable to case-control studies since they have not been calculated separately for cases and controls. We will now consider the derivation of appropriate distributions for cases and controls. We will shift our attention from matching intervals of length l to the entire study period of length L. Let t be the start of the study period. Then the study period consists of the time period $(t, t+L)$. Any individual screened positive in $(t, t+L)$ will be a case in the study. Let T_0 be the last screen prior to t, $M=m$ be the number of screens in the study period, and T_i be the i^{th} screen after t but before $t+L$, $i=1,\ldots,m$. (see diagram below).



Define $U_1=T_1-T_0$, $U_{i+1}=T_{i+1}-T_i$, $i=1,\ldots,m-1$. Thus the U_i's are the intervals between screens ending in the study period. Note that U_1 is a spanning interval and thus we already have its distribution. All we need to do is weight this distribution by $u_1 I$ (where I is the incidence) to obtain the probability of being diagnosed as a case in a spanning interval of length $U_1=u_1$. Thus we will now consider the situations of screen intervals which begin and end within the study period.

IV. Risk of Disease from Intervals Contained in the Study Period.

Consider the joint distribution of two screening times T_i, T_{i+1} in the interval $(t, t+L)$ during which there are $M=m$ screens. For $m \geq 2$,

$$P(T_i \epsilon [t_i, t_i+dt_i],\ T_{i+1} \epsilon [t_{i+1}, t_{i+1}+dt_{i+1}],\ M=m)$$

$$= \frac{\{\mu(t_i-t)\}^{i-1} e^{-\mu(t_i-t)}}{(i-1)!} \cdot \mu dt_i \cdot e^{-\mu(t_{i+1}-t_i)} \cdot \mu dt_{i+1} \cdot \frac{\{\mu(t+L-t_{i+1})\}^{m-i-1} e^{-\mu(t+L-t_{i+1})}}{(m-i-1)!},$$

$$= \mu^m e^{-\mu L} \cdot \frac{(t_i-t)^{i-1}(t+L-t_{i+1})^{m-i-1}}{(i-1)!(m-i-1)!} dt_i dt_{i+1} \quad , \ i=1,2,\ldots,m-1.$$

We wish to obtain the distribution of $U_{i+1}=T_{i+1}-T_i$. Now consider the transformation $(t_i, t_{i+1}) \rightarrow (t_i, u_{i+1})$. The marginal distribution of U_{i+1},

$$f_{U_{i+1},M}(u_{i+1}, m),$$

is obtained by integrating over t_i. Now $t \leq T_i \leq t+L-U_{i+1}$, thus, for $i=1,2,\ldots,m-1$, But,

$$f_{U_{i+1},M}(u_{i+1},m) = \int_{t}^{t+L-u_{i+1}} \frac{(t_i-t)^{i-1}\cdot(t+L-t_i-u_{i+1})^{m-i-1}}{(i-1)!(m-i-1)!}\,dt_i \cdot \mu^m e^{-\mu L}.$$

$$\int_{t}^{t+L-u_{i+1}} \frac{(t_i-t)^{i-1}(t+L-t_i-u_{i+1})^{m-i-1}}{(i-1)!(m-i-1)!}dt_i = \frac{(L-u_{i+1})^{m-1}}{(m-1)!},$$

so that,

$$f_{U_{i+1},M}(u_{i+1},m) = \frac{\mu^m(L-u_{i+1})^{m-1}}{(m-1)!}e^{-\mu L} \quad , \; i=1,2,\ldots,m-1.$$

The above is true for $0 \le u_{i+1} \le L$. For $u_{i+1} > L$, $f_{U+1,M}(u_{i+1},m)=0$.

First notice that the formula for u_{i+1} does not depend explicitly upon i so that the same distribution holds for all u_i, thus for $\forall\ m \ge 2$, we may write,

$$f_{U,M}(u,m) = \frac{\mu^m(L-u)^{m-1}}{(m-1)!}e^{-\mu L}.$$

This (not unexpected) result says that conditional on the number of tests in a diagnostic interval the distribution of each inter-screen interval is the same. Now we must calculate what the probability is that an individual with a screen interval $U=u$, will develop cancer in the study period. Under the null hypothesis this probability is approximately uI, where I is the incidence rate. Conditional on there being $M=m$ tests in the diagnostic interval there will be m-1 test intervals so that the probability that an individual will have cancer diagnosed at the end of an interval of length u in a study period containing m

screens is approximately given by

$$P(D=1,\ U=u,\ M=m) \approx (m-1)\cdot uI\cdot f_{U,M}(u,m)\ ,$$

where the factor (m-1) is included since there are (m-1) intervals. This calculation is approximate because we are ignoring the possibility that disease develops in earlier intervals when calculating the contribution of later intervals. Thus, for $m\geq 2$,

$$P(D=1,\ U=u,\ M=m) \approx uI\cdot\frac{\mu^m(L-u)^{m-1}L}{(m-2)!}e^{-\mu L}\ .$$

In order to obtain $P(D,u)$ we must now sum over m.

$$P(D=1,\ U=u) = \sum_{m=2}^{\infty} uI\cdot\frac{\mu^m(L-u)^{m-1}}{(m-2)!}e^{-\mu L} \qquad (4.20)$$

$$= \begin{cases} \mu^2 u(L-u)Ie^{-\mu u}\ , & 0\leq u<L,\ m\geq 2 \\ 0\ , & otherwise \end{cases}\ .$$

(Notice the sum from $m=2$ since there are no screening intervals within the study period otherwise).

V. Risk of Disease from the Spanning Interval.

We must now calculate the contribution of spanning intervals in which the person is found to have disease at their first test after t. We have calculated the distribution for such tests conditional on one existing (equation 4.18). In this case we are interested in the unconditional probabilities

$$f_U(u) = \begin{cases} \mu^2 u e^{-\mu u} \ , & 0 \le u < L,\ m \ge 1 \\ \mu^2 L e^{-\mu u} \ , & u \ge L,\ m \ge 1 \\ 0 \ , & otherwise \end{cases} . \tag{4.21}$$

VI. Overall Risk of Disease

In order to calculate the overall distribution of cases detected at screening in the study period, $(t, t+L)$, we need only multiply (4.21) by uI and add to (4.20), giving,

$$P(D=1,\ U=u) = \mu^2 u I L e^{-\mu u} \ , \quad u \ge 0 \ . \tag{4.22}$$

But,

$$P(D=1) = \int_0^\infty P(D=1,\ U=u)du = IL \ ,$$

so that the distribution of screening intervals amongst the cases, $P(U=u \mid D=1)$ is,

$$P(U=u \mid D=1) = \mu^2 u e^{-\mu u} \ , \quad u \ge 0,$$

i.e., the distribution of the spanning interval (equation 4.14). Surprisingly, the result doesn't depend on L. Under the null hypothesis the distribution of interval lengths is the same for cases and controls. But we have just found that the distribution of interval lengths for cases is the same as the distribution of spanning intervals, while the distribution of interval lengths for controls is the underlying distribution. The Poisson and uniformity models both agree on these two results.

VII. Inter-subject Variability.

The fixed interval model considered previously has no within subjects variation, while the Poisson model just examined has no between subjects variation. In the case-control framework both models lead to biased estimates of relative risk in that cases tend to have longer intervals than controls under H_0. We can expand the Poisson model to include between subjects effects by mixing the "μ" parameter. That is, assume that there is a distribution $g_M(\mu)$ of μ screening intensities within the population. For the cases we have,

$$P(D=1,\ U=u \mid \mathrm{M}=\mu) = \mu^2 uILe^{-\mu u},$$

from equation (4.22). Thus,

$$P(D=1,\ U=u,\ \mathrm{M}=\mu) = \mu^2 uILe^{-\mu u} g(\mu),$$

$$\Rightarrow P(D=1,\ U=u) = \int \mu^2 uILe^{-\mu u} g(\mu) d\mu,$$

$$\Rightarrow P(D=1) = \int\int \mu^2 uILe^{-\mu u} g(\mu) d\mu du,$$

and thus the conditional distribution is

$$f_{U|D=1}(u) = \frac{\int \mu^2 ue^{-\mu u} g(\mu) d\mu}{\int\int \mu^2 ue^{-\mu u} g(\mu) d\mu du}. \tag{4.23}$$

If $g_M(\mu)$ is chosen to be conjugate to $\mu^2 ue^{-\mu u}$ then $P(u \mid D)$ will have a simple form. The density can be written in exponential form as,

$$\exp[(-\mu)u + \log(u) + \log(\mu^2)]$$

Hence the conjugate prior is a gamma distribution (Cox & Hinkley[22]), say,

$$\frac{\beta^\alpha \mu^{\alpha-1} e^{-\beta\mu}}{\Gamma(\alpha)}$$

Evaluating the numerator of (4.23),

$$\int_0^\infty \mu^2 u e^{-\mu u} \cdot \frac{\beta^\alpha \mu^{\alpha-1} e^{-\beta\mu}}{\Gamma(\alpha)} d\mu = \frac{u\beta^\alpha}{\Gamma(\alpha)} \cdot \int_0^\infty \mu^{\alpha+1} e^{-\mu(u+\beta)} d\mu, \tag{4.24}$$

$$= \frac{u\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+2)}{(u+\beta)^{\alpha+2}},$$

$$= \alpha(\alpha+1) \cdot \beta^\alpha \frac{u}{(u+\beta)^{\alpha+2}}.$$

The denominator evaluates to unity, hence, (4.24) is the conditional density, $f_{U \mid D=1}(u)$.

Next we must consider the controls. We found that for short matching intervals the distribution of screens is approximately $\mu e^{-\mu u}$ (equation 4.19), that is,

$$P(D=0,\ U=u \mid \mathrm{M}=\mu) = \mu e^{-\mu u}.$$

Then,

$$P(D=0,\ U=u,\ \mathrm{M}=\mu) = \mu e^{-\mu u} g_{\mathrm{M}}(\mu),$$

$$p(D=0,\ U=u) = \int \mu e^{-\mu u} g_{\mathrm{M}}(\mu) d\mu,$$

$$P(D=0) = \int\int \mu e^{-\mu u} g_{\mathrm{M}}(\mu) d\mu du,$$

$$f_{U|D=0}(u) = \frac{\int \mu e^{-\mu u} g_{\mathrm{M}}(\mu) d\mu}{\int\int \mu e^{-\mu u} g_{\mathrm{M}}(\mu) d\mu du}.$$

This gives,

$$f_{U|D=0}(u) = \frac{\alpha\beta^{\alpha}}{(u+\beta)^{\alpha+1}} \tag{4.25}$$

Summary and Results

The uniformity and Poisson models give very similar results. First, the distribution of intervals which span a point in time converges to the distribution of intervals starting at any point if the spanning intervals are required to end in a decreasing matching period. Second, the distribution of intervals for cases equals the distribution of spanning intervals under the null hypothesis.

Under the fixed interval model each individual's screening history can be assumed to have an interval which spans a given point in time t. Hence the distribution of spanning intervals is equivalent to the distribution of screening frequencies for sequences of fixed length intervals. In the example, it was assumed that half the population were screened annually and the other half were screened bi-annually. Thus we would find the distribution of spanning intervals to be a half for intervals of length one and two years, the same as the distribution of intervals for the cases. This differs from the distribution of intervals that end at a given time which favoured one year intervals by a ratio of two to one.

We used the B.C. screening program data to investigate the reasonableness of the assumption of uniform starting times. Table 5 presents the distribution of screen intervals which span Dec. 1979 by the starting month for intervals of length 1-5 years.

Table 5. The distribution of screen intervals of lengths 1-5 years which span December, 1979 by month of start of interval.

Years		Dec	Nov	Oct	Sept	Aug	Jul	Jun	May	Apr	Mar	Feb	Jan
1	1979	52	153	200	122	121	120	175	246	180	169	153	169
2	1979	45	68	49	39	23	27	38	46	32	56	53	58
	1978	33	48	48	51	29	28	40	45	46	46	54	49
3	1979	16	24	22	18	12	11	17	17	9	29	9	18
	1978	11	24	17	16	15	9	17	14	19	33	18	16
	1977	7	16	21	12	9	11	12	17	16	16	17	20
4	1979	0	6	4	9	5	8	8	17	13	9	5	8
	1978	5	11	3	3	11	4	7	9	10	17	11	6
	1977	1	8	9	6	3	1	8	5	10	12	9	5
	1976	7	8	14	8	3	3	12	6	4	16	8	6
5	1979	2	9	7	3	5	2	2	2	7	5	2	4
	1978	0	5	5	5	0	1	9	7	6	5	7	2
	1977	2	2	5	3	3	4	7	5	2	8	5	4
	1976	8	4	6	2	2	8	2	2	5	5	6	5
	1975	0	8	7	3	3	3	5	2	9	6	6	12

Except for a slight decrease in screening during vacation periods, the distribution of starting times appears to be quite uniform.

Figure 1 displays the distributions of interval lengths of all intervals which span 1980 (Dec.1979) as well as the distribution of intervals which span 1980 and which end prior to 1981. The third distribution depicted is that of intervals which span 1980 and end prior to 1981, but the frequencies of the intervals have been weighted by the interval lengths.

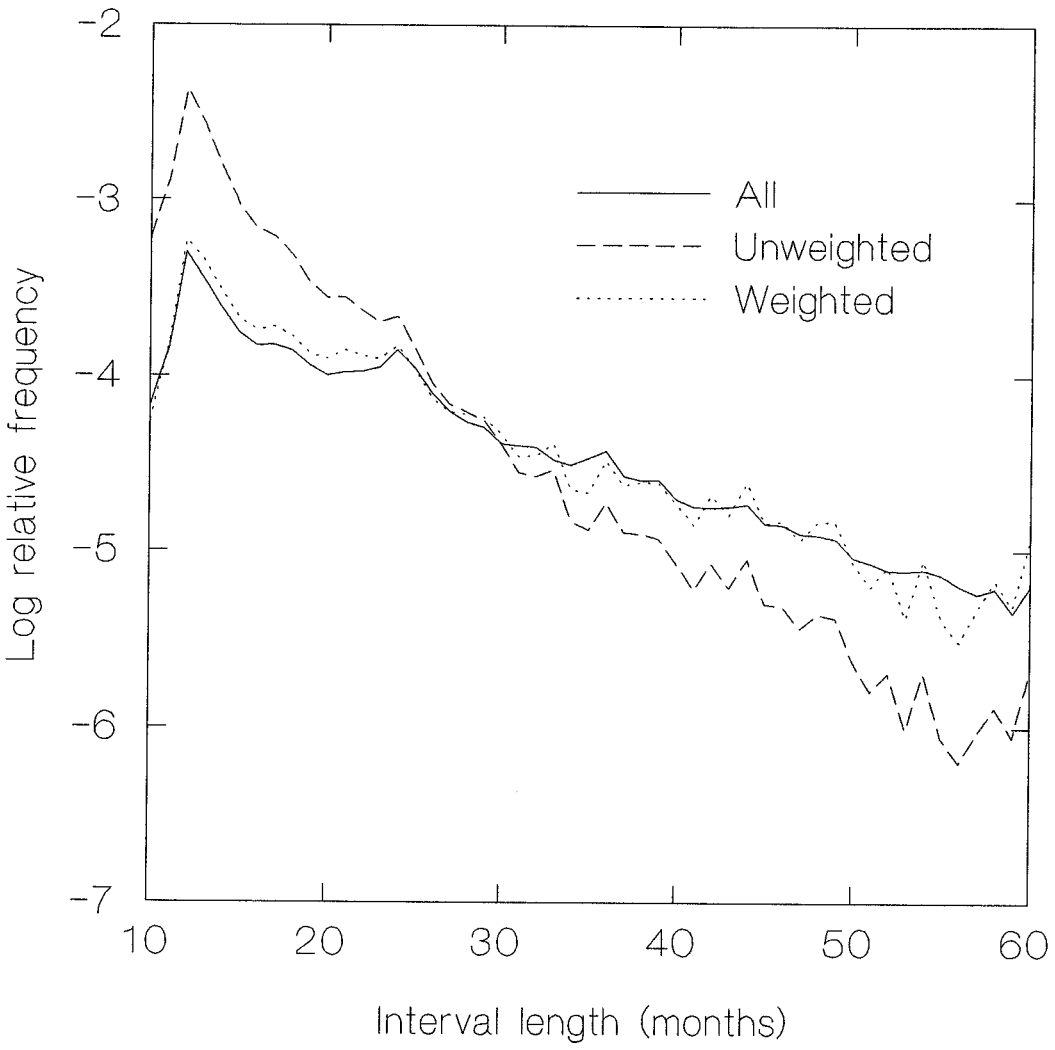


Figure 1. Distribution of intervals spanning 1980 for All intervals and intervals ending before 1981, Unweighted and Weighted by interval length.

The weighted distribution approximates closely the distribution of all spanning intervals. The distribution of unrestricted spanning intervals corresponds to the distribution of "in progress" intervals at a given point in time. If each individual had tests repeated at regular intervals but these intervals varied between individuals then this distribution

would correspond to the distribution of screening frequencies in the population. The distribution of intervals which end in a short time range approximates the distribution of intervals that end at a given point in time. The empirical results displayed in Figure 1 support the relation between the two distributions as predicted under the uniformity assumptions. That is, the probability of an interval ending in a narrow range equals the probability of spanning a point in time weighted inversely by the length of the interval.

Implications of the Poisson model with inter-subject variablility were explored for different parameter values of the conjugate prior representing different degress of variability. The mean and variance of a Gamma distribution, in terms of the parameters α and β, are α/β and α/β^2, respectivley. We chose to fix the mean screening intensity to be the inverse of the unweighted sample mean of the screen interval lengths within the B.C. cervical screening program data set (to be described below). The mean interval length for 668,751 intervals was 27.10 months with a standard deviation of 27.39 months. Values for α and β selected to give a mean of 1/27.1 and coefficients of variation of 30%, 60%, and 100% to represent low, medium, and high variability respectively. The principle outcome of interest is the resulting odds ratio for various categories of interval length. To parallel the analyses to follow, we chose interval categories: 10-18 months, 19-30 months, 31-42 months, 43-54 months, 55-120 months, and >120 months. The odds ratios of interval category "exposure", (e.g. exposure to a 10-18 month screen interval), relative to the index category, >120 months, are given by,

$$\frac{P(U\in[10,\ 18]\ |\ D=1)\ /\ P(U\in(120,\ \infty)\ |\ D=1)}{P(U\in[10,\ 18]\ |\ D=0)\ /\ P(U\in(120,\ \infty)\ |\ D=0)}$$

The probabilities were obtained by integrating the appropriate conditional density over the specified range. The results are presented in Table 6.

Table 6. Theoretical odds ratio of interval length relative to >120 months, cases vs. controls, by amount of inter-subject screening intensity variability in Poisson model.

Interval length	Variability (coefficient of variation)		
	low (30%)	medium (60%)	high (100%)
10-18 months	0.125	0.280	0.368
19-30 months	0.212	0.418	0.516
31-42 months	0.307	0.535	0.629
43-54 months	0.396	0.622	0.705
55-120 months	0.583	0.763	0.818

Evidently lower inter-subject variability in screening intensity results in more pronounced effects of interval length on prevalence odds ratios.

Chapter 5

Data Analysis

B.C. has a centralised cytology screening program which has been in operation since 1949. Until the early sixties, it was more of a diagnostic support service than a screening program. The proportion of women over 20 ever screened was about 3% in 1955[1]. With the widespread use of oral contraceptives in the early sixties, the Pap test, which was applied in conjunction with the dispensing of oral contraceptives, became increasingly less of a diagnostic tool and more of a screening test. By 1962 it was estimated that 53% of women over 20 had ever been screened, and by 1969 this figure rose to 78%[1]. The samples are obtained by general practitioners and gynaecologists for the most part, and then sent to a central laboratory where they are interpreted. Patients are assigned identification numbers which their physicians are supposed to provide along with the sample for patient identification, but name and date of birth are also recorded. The test results and other information are entered into a centralized cytology computer file. The result, along with recommendations for further care, is returned to the referring physician who is responsible for the care of the patient. Physicians are responsible for advising appropriate screening intervals for their patients. If, for example, a test is abnormal and not merely benign atypia, physicians will be sent a

reminder for a repeat test if one is not obtained within four months. Also, for cases with histories of severe dysplasia or carcinoma *in situ*, reminders are automatically sent annually.

D.A. Boyes, B. Morrison, and colleagues[1], undertook a cohort study with data from the British Columbia Screening Program covering the years 1949-1969. Their objective was to provide estimates of prevalence and incidence rates of dysplasia or worse and carcinoma *in situ* or worse. Two cohorts were selected to cover as wide a range of age as possible as well as providing some overlap. "The records of all women who had been born in the years 1914-1918, and 1929-1933, and who had had 1 or more cervical tests [prior to 1969], were pulled from the identity files of the British Columbia Central Cytology Laboratory"[1] (52,452 and 66,701 women respectively). Extensive record linkage procedures were carried out in order to minimize duplications. The following information was extracted from the data base for every qualifying women:

(1) Identifying information, i.e. the patient's surname (first 12 characters), first name (first 8 characters), second initial, month and year of birth, and husband's first name (first 4 characters);

(2) The month, year and cytological class of every smear taken up to the end of 1969 and for women with subsequent positive histological findings, the original cytological class as assigned when the smear was first read, as well as a reviewed class based upon a re-

examination of the specimens;

(3) All consequent radiation and surgical procedures, including cervical biopsy and hysterectomy. Hysterectomies performed for reasons that were not the consequence of findings obtained from screening could not be completely documented unless a woman came for a repeat smear following hysterectomy.

(4) Histological diagnoses based on biopsies or other surgical procedures.

Follow-up on the original two cohorts has been updated to 1992 by Morrison[21]. This provides a longer period of observation for each woman as well as a greater overlap in age for the two cohorts (30 years instead of 5). The original study involved 121,722 woman. Successful linkage was achieved for 43% providing updated data for 71,236 women. There were two reasons why linkage failed in about half the women. (1) Files which had a history of either all negative, no histology, no test in past 7 years, or death between 1976 and 1985 were removed from the archives. (2) The linkage method was conservative.

The "raw" data were prepared for analysis as follows: First, the data consisting of PAP test, histology, and death records were sorted chronologically within subjects. The records were then processed sequentially. A pair of records was defined to constitute a screen interval if the starting record had class 1, the ending record was a Pap test record,

and the interval was at least 10 months. This operational definition of screening intervals was designed to achieve two objectives. First, we wanted to examine the risk of disease associated with routine screening, as opposed to diagnostic testing which is often times done when the presence of disease is suspected. And since we didn't have any information other than the times and results of tests, we had to resort to the method of excluding intervals under 10 months. We decided to exclude intervals that started with an abnormal test result for the same reason. The excluded intervals create "gaps" in an individual's screening history. Although the gaps are not studied as screen intervals, they do play a role in the estimation of incidence rates and in the formation of covariate classes based on conditions preceding screen intervals.

Any records following either diagnosis of CIS or worse or a hysterectomy were excluded.

Each screen interval was assigned to a covariate class to contribute towards the comparison of cases and non-cases. For the cases this interval was the last interval before diagnosis. So each valid screen interval is included as a "last" interval and contributes to a covariate class defined by the following factors:

(1) Cohort: born 1914-1918 or 1929-1933,

(2) Period: the time period when the interval ended was categorized as: pre- 1963, 1963-

1975, 1976-1992. These cut points were chosen because the three periods were believed to differ with respect to sample characteristics and/or screening practices. Prior to 1963 screening tended to serve primarily as a diagnostic tool. At about 1963 the use of oral contraceptives became more prevalent and since Pap tests were frequently performed in conjunction with the dispensing of oral contraceptives the test became more of a screening instrument. Prior to the mid 70s the diagnosis and treatment of abnormalities was quite a serious procedure, so it was common practice to wait until the disease had reached an advanced stage before taking action. However with the introduction of the colposcope (an instrument which allows the visual inspection of the tissue) in the mid 70s, treatment was applied more readily.

(3) Interval: screen interval lengths were broken into categories: 10-18, 19-30, 31-42, 43-54, 55-119, 120+ months, the idea being to approximate 1,2,3,4, and 5-10 year intervals. This category is referred to as the "last interval" category since cases were assigned to the "last" interval.

(4) Previous: the combined lengths of the two screen intervals preceding the "last" interval were grouped in the following categories: 10-36, 37-60, 61-84, 85-119, 120+ months, *no preceding intervals*, and *one preceding interval.* The reason for looking at preceding intervals is to address the fact of false negative tests, i.e. tests which report no abnormalities when in fact disease is present. Thus an individual who tests negative twice is less likely to have disease than one who has tested negative only once. It may

be useful to be able to estimate the risk of disease associated with various patterns of screening intervals. The choice of categories for the combined length of the two preceding intervals is intended to group individuals who were tested at fixed intervals of one, two, three, and four years. It also identifies individuals who had one or no preceding screens.

(5) Preceding gap abnormality: three categories of abnormality occurring in the "gap", if any, preceding the "last" interval: one for intervals with no preceding gap (either the "last" interval began with the termination of a screen interval with a class 1 test result or it was the individual's first record); a second category for "minor" signs of abnormality including class 2 or class 9 (inadequate sample) test results or class 1, 2, or 9 less than 10 months apart; and a third category for "major" signs of abnormality including class 3 or 4 test results or any histology record. The inclusion of this factor was motivated by an attempt to control what is probably a very important predictor of disease, namely previous abnormalities.

The date of diagnosis for cases with CIS or worse was defined to be the endpoint of the last screen interval preceding diagnosis. The rationale for this definition was that tests done subsequent to the last screen interval were probably done for diagnostic purposes rather than routine screening. The exact date of onset of disease cannot be identified with much accuracy, so the decision is somewhat arbitrary. The total time-at-risk contributed by an individual extends from the start of the first screen interval to the end

of the last record or until diagnosis or hysterectomy. It is partitioned into segments according to starting points of screen intervals. Gaps are assigned to the preceding screen interval. The segments are then assigned to covariate classes according to the covariate pattern of the screen interval defining the segment. Period is defined by the endpoints of the screen intervals. This method of assigning time at risk is method B of Boyes et. al. and follows the principle that time at risk is assigned to the category to which the case would have been assigned had disease developed at that time.

There were a total of 1198 cases of CIS or worse among 1.1 million records from roughly 120,000 subjects. The numbers of cases and incidence rate estimates by covariate class are given in Appendix 1. There appears to be a trend for incidence rates to decrease with increasing interval length.

A Poisson regression analysis[30] was performed on the incidence data, predicting the log of the number of cases from cohort, period, length of last interval, combined length of two intervals preceding the "last" interval, and the degree of abnormality in the gap preceding the "last" interval. Log time-at-risk was included as the offset. Poisson regression fits the log of the estimated incidence rates to a linear combination of the covariates which has the effect of treating the covariates as having multiplicative effects on incidence rates. The covariates, being categorical, are represented by terms for all but one of the category levels - the index category. Treatment contrasts were used to compare the effect of each level with the effect of the index category. The combination

of all index categories is represented by the intercept. Let n represent the number of cases, t the aggregated time-at-risk and **x** the vector of factor categories and an intercept, then the model can be expressed as,

$$\log\{E(n)\} = \log(t) + \beta' x$$

The results of the model fit are presented in Table 7.

Table 7. Poisson regression of incidence rates on length of "last" interval, combined length of two preceding intervals, and abnormality in gap preceding the "last" interval.

Parameter		Coef	s.e.	t value	risk
Intercept		-11.1	0.330	-33.7	
Cohort	born 1929-33	0.543	0.256	2.12	
Period	1963-1975	0.192	0.200	0.958	
	1976-1992	0.074	0.221	0.336	
Period*Cohort	period*cohort1	0.094	0.268	0.352	
	period*cohort2	-0.426	0.281	-1.52	
Length of last interval	55-120 months	0.273	0.175	1.56	1.31
	43-54 months	0.581	0.168	3.46	1.79
	31-42 months	0.890	0.166	5.35	2.44
	19-30 months	0.966	0.161	6.02	2.63
	10-18 months	1.31	0.159	8.22	3.71
Combined length of two intervals preceding the last	85-120 months	0.217	0.235	0.921	1.24
	61-84 months	-0.051	0.223	-0.230	0.950
	37-60 months	-0.478	0.211	-2.26	0.620
	10-36 months	-0.605	0.210	-2.88	0.546
	one previous	0.259	0.214	1.21	1.30
	no previous	0.620	0.212	2.92	1.86
Preceding gap abnormality	minor	0.507	0.067	7.59	1.66
	major	1.23	0.217	5.70	3.42

Index categories: >120 months for last and preceding two intervals and no preceding gap for preceding gap abnormality.

The inclusion of an interaction term for cohort and period significantly improved the fit

and thus was included in the model. No other interactions were significant. The model was also fit without an offset, and including log of the time at risk whose coefficient was found to be 0.943 with a standard error of 0.052. This implies that the data are consistent with assumed relation between incidence and time-at-risk. Visual inspection of residuals plotted against fitted values and by factor did not indicate any blatant signs of misfit. The dispersion parameter was estimated to be 1.1 using formula (6.4) from McCullagh & Nelder[30] which is given by,

$$\tilde{\sigma}^2 = X^2/(n-p) = \sum_i \frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i}/(n-p).$$

where n is the number of cells with non-zero time-at-risk, p is the number of parameters in the model, y_i is the observed count in cell i, $\hat{\mu}_i$ is the model predicted count for cell i. A dispersion value greater than one suggests that there is 'over-dispersion' in the data[30]. The residual deviance was 406.3 on 552 degrees of freedom.

The rates presented in Table 7 are derived from taking exponents of the coefficients. Since "treatment" contrasts were used (i.e., the coefficient of the index factor category was set equal to zero), taking the exponents of other categories provides estimates of incidence rates relative to the index category. The index category was taken to be >120 years for both interval length factors and *no preceding abnormality* for the preceding gap abnormality factor. The baseline incidence rate for the covariate class defined by all index categories is given by the exponent of the intercept which equals 0.181 cases per

1000 person-years (note that the exponent of the intercept must be multiplied by 12,000 to transform the units from person-months to 1000 person-years). The degrees of freedom can be taken to be 552[30]. Thus the t values can be compared with 1.96 for significance at the .05 level. All but one of the interval length levels were significant as well as the two shortest combined preceding interval lengths, the level representing no previous screens, and both levels of the preceding condition factor. It would seem that length of screen interval is associated with reduced risk of disease, but having had a couple of screens in the recent past is beneficial if the outcome is negative.

The Poisson model can be used to estimate incident rates for joint combinations of factors. If we match up categories from the combined lengths of the preceding two intervals with the length of the current interval, we can get an approximation of the risk of disease with annual, bi-annual, tri-annual, and quadra-annual screening patterns. Since if an interval length is to be recommended, it is assumed that it will be followed on a regular basis. Thus, for example, the interval length category of 19-30 months will be combined with the category of 37-60 months for the combined length of the two previous intervals. Together they represent the risk of a bi-annual screening pattern. The incidence rate relative to the combination of baseline categories is estimated by the product of the exponents of the respective coefficients. The estimated rates relative to baseline are: 2.02 for annual screeners, 1.63 for bi-annuals, 2.31 for tri-annuals, and 2.22 for individuals who are screened every four years. When the effect of previous screens is added to the effect of the "last" screen the two effects tend to dilute each

other. But the effect of the "last" evidently wins out since the incidence rate is still twice the rate of the reference group which consists of individuals whose "last" interval was greater than 10 years and whose combined lengths of the previous two intervals is also greater than 10 years.

Incidence rates are based on numbers of cases occurring in units of time. We also examined the proportion of intervals in each covariate class which resulted in diagnosis of disease. This gives an index of the prevalence of disease in each covariate class. The prevalence rates are given in Appendix 2. Here the reverse trend of rate with respect to interval length is observed. We modelled the prevalence rates with logistic regression, using the same set of covariates as in the Poisson model. The results are summarized in Table 8.

Treatment contrasts were again used so the exponents of the coefficients give estimates for the odds ratios of each of the factor levels with respect to the index levels. The results are the same as the Poisson regression results except for the effect of interval length which in this case indicates decreasing risk with decreasing length rather than the other way around. But this does not necessarily imply that individuals are better off with shorter intervals because you have to make it through relatively more intervals when they are shorter. The proper index for comparison is the risk per unit time. However this does not quite equal incidence rates as they have been computed here, since simply dividing the prevalence rate for an interval by the length of the interval does not account

Table 8. Logistic regression of prevalence cases on length of "last" interval, combined length of two preceding intervals, and abnormality in gap preceding the "last" interval.

Parameter		Coef	s.e.	t value	risk
Intercept		-5.98	0.329	-18.2	
Cohort	born 1929-33	0.526	0.257	2.05	
Period	1963-1975	0.210	0.201	1.04	
	1976-1992	0.120	0.222	0.539	
Cohort*Period	cohort*period1	0.131	0.269	0.488	
	cohort*period2	-0.407	0.282	-1.45	
Length of last interval	55-120 months	-0.430	0.175	-2.46	0.651
	43-54 months	-0.601	0.167	-3.59	0.548
	31-42 months	-0.652	0.165	-3.94	0.521
	19-30 months	-0.955	0.159	-5.99	0.385
	10-18 months	-1.08	0.158	-6.84	0.340
Combined length of two intervals preceding the last	85-120 months	0.230	0.236	0.973	1.26
	61-84 months	-0.040	0.224	-0.177	0.961
	37-60 months	-0.470	0.212	-2.21	0.625
	10-36 months	-0.613	0.211	-2.91	0.542
	one previous	0.291	0.215	1.35	1.34
	no previous	0.658	0.213	3.09	1.93
Preceding gap abnormality	minor	0.554	0.067	8.29	1.74
	major	1.28	0.219	5.88	3.60

Index categories: >120 months for last and preceding two intervals and no preceding gap for preceding gap abnormality.

for the time between intervals, the "gaps" during which diagnostic testing is presumably occurring.

By equation (3.2) the ratio of prevalence odds to incidence rate equals the average duration of disease. We computed model estimates of prevalence odds and incidence rates for each level of the "last" interval category. We then divided the prevalence odds by the corresponding incidence rate to obtain an estimate of the average duration of disease for each "last" interval category. The results are presented in Table 9.

Table 9. Model estimated incidence rates, prevalence odds and the ratio of prevalence odds to incidence rates by "last" interval length.

"Last" interval length	Incidence rate (/1000 person-years)	Prevalence rate (/1000 persons)	Prevalence/ Incidence (years)
10-18 months	0.673	0.860	1.28
19-30 months	0.477	0.974	1.45
31-42 months	0.442	1.32	2.99
43-54 months	0.325	1.39	4.28
55-120 months	0.238	1.65	6.93
>120 months	0.181	2.53	14.0

The categories correspond roughly to intervals of length 1, 2, 3, 4, 5-10, and >10 years. Clearly the estimated duration of disease is related to "last" interval length. One would expect the duration of disease to be longer on average if diagnosed after a long screen interval rather than a short screen interval. However, the average duration estimates are unrealistic in that they seem to correspond quite closely with the average interval length within each category, which would imply that on average disease occurred shortly after the start of the screen interval. Perhaps the duration is inflated by the occurrence of false negative tests.

The assumption of constant incidence overlooks one important factor, namely the stage of disease between normality and CIS, namely dysplasia. Assuming a negative test to

be accurate, an individual cannot develop CIS without first going through a period of dysplasia. There is a period of grace, namely the sojourn time for dysplasia, during which CIS cannot occur since dysplasia must first run its course. Hence the interval over which disease can develop is shorter than the nominal interval. This would have the effect of deflating the ratios of incidence rates to prevalence odds ratios, which may account for the observed results.

We also simulated the methodology of the case-controls studies discussed earlier. Each of the cases was matched with seven controls on the basis of year of birth and to within five months of "date of diagnosis". That is, controls were required to have had a screen interval end within five months of the endpoint of the last screen interval before diagnosis of the matched case. The data were then subjected to a conditional likelihood logistic regression analysis using the PECAN package[31]. The same covariates were used excluding cohort and period since these were controlled by design. The results are presented in Table 10.

The same pattern of results are observed as for the unconditional logistic regression model, although the two models are not identical. The case-control study matched by year of birth and date of diagnosis to within 5 months, while the cohort study merely controlled for cohort and period of diagnosis to within 1-2 decades. The risk estimates are somewhat less extreme for interval length and more extreme for the combined length

Table 10. Conditional likelihood logistic regression of a simulated matched case-control study.

Parameter		Coef	s.e.	z score	risk
Length of last interval	55-120 months	-0.744	0.228	-3.27	0.475
	43-54 months	-0.909	0.219	-4.14	0.403
	31-42 months	-0.954	0.220	-4.35	0.385
	19-30 months	-1.31	0.213	-6.16	0.269
	10-18 months	-1.42	0.212	-6.68	0.242
Combined length of two intervals preceding the last	85-120 months	0.014	0.267	0.054	1.02
	61-84 months	-0.157	0.250	-0.628	0.855
	37-60 months	-0.703	0.237	-2.96	0.495
	10-36 months	-0.858	0.235	-3.65	0.424
	one previous	0.044	0.243	-0.180	0.957
	no previous	0.292	0.244	1.19	1.34
Preceding gap abnormality	minor	0.544	0.076	7.20	1.72
	major	1.20	0.272	4.40	3.31

Index categories: >120 months for last and preceding two intervals and no preceding gap for preceding gap abnormality.

of preceding screen intervals. The standard errors also tend to be slightly larger. Although the two methods of modelling prevalence rates differ to some extent, they both suggest the same conclusion, namely that interval length is directly related to risk of disease, contrary to results of the analysis of incidence rates.

We have seen that prevalence rates, as given by unconditional logistic regression with cohort data or conditional likelihood regression with case-control data, are inflated relative to incidence rates by a factor roughly equal to the average length of duration of disease. Thus, crude estimates of incidence rates could be obtained from prevalence rates by adjusting for the likely relation between the factor and disease duration. Another possibility would be to weight the sampling of controls, in case-control studies, to adjust for the effect of interval length as was done in Figure 1. This has the effect of

changing the distribution of interval lengths from that of intervals starting or terminating at a point in time to that of intervals spanning a point, which, as we have seen, under the uniform and Poisson models of screening, corresponds to the distribution of interval lengths for cases under the null hypothesis. This method may have the advantage, over simply correcting the obtained prevalence odds, of simultaneously correcting odds ratios obtained for factors which may be associated with interval length.

The suggested method was examined in another simulated, matched, case-control study using a weighted sample of controls. The criteria for control selection were the same as before, only this time the probability of sampling any given control from those eligible was weighted inversely by the length of the "last" interval (to the nearest year). The results are presented in Table 11.

Table 11. Conditional likelihood logistic regression of a simulated matched case-control study with sampling of controls weighted by length of "last" interval.

Parameter		Coef.	s.e.	z score	risk
Length of last interval	55-120 months	-0.062	0.187	-0.335	0.940
	43-54 months	1.15	0.186	6.21	3.17
	31-42 months	0.834	0.183	4.56	2.30
	19-30 months	0.977	0.177	5.51	2.66
	10-18 months	1.70	0.179	9.50	5.48
Combined length of two intervals preceding the last	85-120 months	0.146	0.281	0.520	1.16
	61-84 months	-0.216	0.265	-0.818	0.806
	37-60 months	-0.671	0.251	-2.67	0.511
	10-36 months	-0.836	0.250	-3.35	0.433
	one previous	-0.031	0.256	-0.122	0.969
	no previous	0.437	0.258	1.69	1.55
Preceding gap abnormality	minor	0.515	0.076	6.74	1.67
	major	1.16	0.293	3.95	3.18

Index categories: >120 months for last and preceding two intervals and no preceding gap for preceding gap abnormality.

Weighting the sampling of controls by length of "last" interval succeeded in reversing the direction of trend in the relation between length of "last" interval and relative risk. The resulting risk estimates are not too far off those produced by the Poisson regression model, although the estimate for the 43-54 month category appears to be a little high. Strangely, the standard errors are also closer those of the Poisson model than the ones produced by the unweighted sample matched case-control analysis.

The ratios of unweighted sample (prevalence) to weighted sample ("incidence") odds ratios is 0.044, 0.102, 0.156, 0.127, and 0.505 for "last" interval categories corresponding to 1,2,3,4, and 5-9 years respectively. The relative sizes roughly correspond to the relative lengths of the intervals except for the "4" category.

Chapter 6

Conclusion

The analysis of screen detected disease involves a number of issues with respect to the design and analysis of cohort and case-control studies. Non-standard methods are required to estimate incidence rates for time-dependent factors such as interval length. Case-control studies of screen detected disease produce estimates of prevalence odds ratios rather than incidence risk ratios. This can be misleading when the exposure variable is related to duration of disease. In the case of screening interval length, since CIS is a long lasting disease, it is not surprising that higher prevalence rates are observed among individuals with longer screening intervals. But higher prevalence rates do not imply higher incidence rates, since after the completion of one screen interval another one begins with, perhaps, the same risk of disease.

To examine the effect of a method of sampling controls in a popular case-control paradigm, whereby controls are required to have had a screen near the time of diagnosis of a case, theoretical screening models were considered. All models imply that the distributions of interval lengths are different for cases and controls under the null hypothesis of no effect of screening. The distribution of interval lengths for sampled

controls converges to the underlying distribution, whereas the distribution of interval lengths for cases is the same as the distribution of spanning intervals, favouring longer intervals.

Incidence rates evidently cannot be accurately inferred from prevalence rates unless the duration of disease is known. Brookmeyer, Day, and Moss[32] have proposed a method of estimating the duration of disease along with false negative rates. Unfortunately, the method presupposes that regression does not occur, and this would certainly not be the case for CIS.

With respect to the specific findings of this study, it would appear that frequent screening does not provide protection against the development of CIS. It is beneficial to have had negative screens in the past, but the incidence of diagnosed CIS is higher for shorter screening intervals than for longer ones. One possible explanation for this paradoxical result is that the natural regression of the disease may result in some cases, which arise and then regress during long screen intervals, going undetected. The time that these individuals actually had the disease is being erroneously applied to the time-at-risk, deflating the incidence rate. And even more importantly, an incidence case is going uncounted. Such errors would be less likely to occur under a frequent screening pattern. However, frequent screening should result in some avoidance of disease in that precursor stages of disease, in this case dysplasia, are detected and treated. Thus frequent screeners should be disease free altogether, except for rapidly developing subtypes.

Another possible explanation for the observation of higher incidence rates in shorter intervals is that individuals who are at risk are more likely to go for frequent tests. We tried to control for this somewhat by controlling for degree of abnormality occurring in the gap preceding the "last" interval. However, it is generally believed that high risk individuals are less likely to be screened frequently.

Although recency of previous tests provides some protection against disease, it does not appear to be enough to counter the effect of interval length. It would appear that the only logical conclusion would be that the best protection against the diagnosis of CIS is infrequent screening.

A shortcoming of the present study is the lack of control or information about the circumstances surrounding the decision to be tested. The possibility of confounding variables is always a concern with observational studies. Even though the present study casts some doubt on the benefit of frequent screening for the prevention of diagnosis of CIS, as opposed to invasive cancer, a proper randomized control trial may still be considered unethical because of the possible implications for more advanced stages of disease.

Bibliography

[1] Boyes D.A., Morrison E.G., Knox E.G., Draper G.J., Miller A.B., *A cohort study of cervical cancer screening in British Columbia.* Clinical & Investigative Medicine, **1982**; 5:1-29.

[2] Ellman R., Indications for colposcopy from a UK viewpoint, in Miller A.B., Chamberlain J., Day N.E., Hakama M., Prorak P.C. (Eds) *Cancer Screening*, **1991**. Cambridge University Press.

[3] Day N.E., *Screening for cancer of the cervix*, Journal of Epidemiology and Community Health, **1989**, 43:103-106.

[4] Anderson G.H., Boyes D.A., Benedet J.L., et. al. *Organisation and results of the cervical cytology screening programme in British Columbia, 1955-85*, British Medical Journal, **1988**, 296:975-978.

[5] Task Force Appointed by the Conference of Deputy Ministers of Health: *Cervical cancer screening programs*, Canadian Medical Association Journal, **1976**, 114:1003-1033.

[6] Parkins in Hakama M., Miller A.B., Day N.E., (Eds) *Screening for Cancer of the Uterine Cervix.* **1986**. IARC Scientific Publications, No. 76: Lyon.

[7] La Vecchia C., Franceschi S., Decarli A., et. al. *"PAP" smear and the risk of cervical neoplasia: quantitative estimate from a case-control study*, The Lancet, **1984**, Oct:779-782.

[8] Knox G. *Case-control studies of screening procedures*, Public Health, **1991**, 105:55-61.

[9] Chronic Disease Reports: *Deaths from cervical cancer - United States 1984-1986.* MMWR, **1989**, 38:650-659.

[10] Miller A.B., Knight J., Narod S. *The natural history of cancer of the cervix, and the implication for screening policy*, in ...

[11] Hutchinson M.L., Agarwal P., Denault T. et. al. *A new look at cervical cytology*, Acta Cytologica, **1992**, 36:499-504.

[12] Hakama M., Miller A.B., Day N.E. (Eds) *Screening for Cancer of the Uterine Cervix*. **1986**. IARC Scientific Publications, No. 76: Lyon.

[13] Magnus and Langmark in Hakama M., Miller A.B., Day N.E. (Eds) *Screening for Cancer of the Uterine Cervix*. **1986**. IARC Scientific Publications, No. 76: Lyon.

[14] Berrino in Hakama M., Miller A.B., Day N.E. (Eds) *Screening for Cancer of the Uterine Cervix*. **1986**. IARC Scientific Publications, No. 76: Lyon.

[15] Clark E.A., Anderson T.W., Does screening by 'PAP' smears help prevent cervical cancer? *The Lancet*, **1979**, July:1-4

[16] Berrino F., *Papanicolaou smears and risk of cervical neoplasia*, The Lancet, **1984**, Nov:1099-1100.

[17] MacGregor J.E., Moss S.M., Parkin D.M., Day N.E., *A case-control study of cervical cancer screening in north east Scotland*, British Medical Journal, **1985**, 290:1543-1546

[18] Rothman K.J., *Modern Epidemiology*, **1986**, Little Brown and Company:Toronto.

[19] van Oortmarssen G.J., Habbema J.D.F., in Hakama M., Miller A.B., Day N.E., (Eds.) *Screening for Cancer of the Uterine Cervix*. **1986**. IARC Scientific Publications, No. 76: Lyon.

[20] Choi N.W., Nelson N.A. in Hakama M., Miller A.B., Day N.E. (Eds) *Screening for Cancer of the Uterine Cervix*. **1986**. IARC Scientific Publications, No.76: Lyon.

[21] Morrison B.J., Coldman A.J., Boyes D.A., Anderson G.H., Forty year follow-up of cervical screening: the cohort study of British Columbia, **1994.** Unpublished manuscript.

[22] Habbema J.D.F. in Hakama M., Miller A.B., Day N.E. (Eds) *Screening for Cancer of the Uterine Cervix*. **1986**. IARC Scientific Publications, No.76: Lyon.

[23] Cox D.R., Hinkley D.V. *Theoretical Statistics*, Chapman and Hall, New York, **1974**

[24] Armitage P., Berry G. *Statistical Methods in Medical Research*, **1987**, Blackwell Scientific Publications:London.

[25] Sheehe P.R. Dynamic risk analysis in retrospective matched pair studies of disease. *Biometrics* **1962**; 18:323-341.

[26] Miettinen O.S. Estimability and estimation in case-referent studies. *Am. J. Epidemiol.* **1976**; 103:226-235.

[27] Breslow N.E., Day N.E., *Statistical Methods in Cancer Research*, vol.1 **1980**, vol.2 **1987** International Agency for Research on Cancer: Lyon.

[28] Elandt-Johnson R. *Definition of rates: some remarks on their use and misuse.* American Journal Of Epidemiology, **1975**; 102:267-271.

[29] Sasco A.J., Day N.E., Walter S.D., *Case-control studies for the evaluation of screening*, Journal of Chronic Disease, **1986**, 39:399-405.

[30] McCullagh P., Nelder J.A., *Generalized linear models*, (2nd Ed.) Chapman and Hall: New York, **1989**.

[31] Storer B.E., Wacholder S., Breslow N.E., Maximum likelihood fitting of general relative risk models to stratified data. *Applied Statistics*. **1983**, 32:172-81.

[32] Brookmeyer R., Day N.E., Moss S. Case-control studies for estimation of the natural history of preclinical disease from screening data, *Statistics in Medicine*, **1986**, 5:127-138.

Appendix 1

Number of cases and incidence rates per thousand person-years by:

- Cohort (1=born 1914-1918, 2= born 1929-1933)
- Period (1=pre-1963, 2= 1963-1975, 3=1976-1992)
- Degree of abnormality in Gap preceding last screen (1=no gap, 2=minor-<10 months or class 2 or 9, 3=major-class 3, 4 or histology)
- Length of last screen interval,
- Combined length of two screens prior to last.
- Cells with no time-at-risk indicated with "-"

Cohort 1	Length of last screen interval (months)											
Period 1	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	9	1.76	4	0.74	3	0.80	1	0.25	0	0.00	0	0.00
1 prev	2	0.74	1	0.56	2	2.50	0	0.00	0	0.00	0	-
10-36	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
37-60	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
61-84	0	0.00	0	0.00	1	53.33	0	0.00	0	-	0	-
85-120	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
>120	0	0.00	0	-	0	-	0	-	0	-	0	-
Gap 2												
No prev	6	6.09	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
10-36	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-
37-60	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-
61-84	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
85-120	0	0.00	0	-	0	0.00	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-
Gap 3												
No prev	0	0.00	0	0.00	0	0.00	0	-	0	0.00	0	-
1 prev	0	0.00	0	-	0	-	0	-	0	-	0	-
10-36	0	-	0	-	0	-	0	-	0	-	0	-
37-60	0	-	0	0.00	0	-	0	-	0	-	0	-
61-84	0	-	0	-	0	-	0	-	0	-	0	-
85-120	0	-	0	-	0	-	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-

Cohort 1	Length of last screen interval (months)											
Period 2	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	11	1.42	12	0.98	11	0.96	14	0.75	7	0.32	6	0.70
1 prev	10	0.85	12	0.95	3	0.35	5	0.52	6	0.87	2	1.19
10-36	13	0.35	5	0.34	1	0.20	1	0.28	0	0.00	0	0.00
37-60	13	0.82	2	0.15	1	0.17	2	0.43	0	0.00	0	0.00
61-84	3	0.80	2	0.46	1	0.42	0	0.00	1	0.99	0	0.00

85-120	2	1.14	1	0.56	2	1.89	2	2.44	0	0.00	0	0.00
>120	1	1.65	1	1.79	1	3.16	0	0.00	0	0.00	0	-
Gap 2												
No prev	5	2.21	2	1.00	3	2.22	3	1.42	0	0.00	1	1.20
1 prev	2	0.59	1	0.40	0	0.00	3	1.92	2	1.88	0	0.00
10-36	6	0.58	4	0.94	1	0.84	0	0.00	1	2.01	0	0.00
37-60	5	1.13	2	0.71	0	0.00	0	0.00	0	0.00	0	0.00
61-84	3	2.80	1	1.07	0	0.00	0	0.00	0	0.00	0	0.00
85-120	0	0.00	1	2.68	0	0.00	0	0.00	0	0.00	0	-
>120	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
Gap 3												
No prev	2	27.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
1 prev	1	12.24	1	26.49	0	0.00	0	0.00	0	-	0	-
10-36	1	5.77	0	0.00	0	0.00	0	0.00	0	0.00	0	-
37-60	1	12.30	1	42.86	0	0.00	0	0.00	0	0.00	0	-
61-84	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
85-120	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
>120	0	0.00	0	0.00	0	-	0	-	0	-	0	-

Cohort 1	Length of last screen interval (months)											
Period 3	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	5	2.39	6	0.31
1 prev	0	0.00	1	1.55	4	4.91	0	0.00	0	0.00	4	0.34
10-36	5	0.20	5	0.33	0	0.00	1	0.13	1	0.12	2	0.26
37-60	5	0.44	3	0.21	2	0.20	3	0.24	2	0.15	0	0.00
61-84	4	1.25	7	1.27	2	0.40	2	0.28	3	0.38	1	0.22
85-120	1	0.52	1	0.32	2	0.65	1	0.21	2	0.33	1	0.33
>120	0	0.00	4	1.61	0	0.00	0	0.00	2	0.41	0	0.00
Gap 2												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.36
1 prev	0	0.00	0	0.00	0	0.00	1	2.83	0	0.00	1	0.52
10-36	3	0.52	4	1.14	2	1.35	0	0.00	0	0.00	0	0.00
37-60	1	0.33	1	0.38	1	0.66	1	0.60	2	1.04	1	0.78
61-84	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
85-120	0	0.00	0	0.00	1	2.38	0	0.00	1	1.62	0	0.00
>120	1	2.44	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Gap 3												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
10-36	1	1.80	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
37-60	1	2.43	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
61-84	1	8.06	0	0.00	0	0.00	0	0.00	0	0.00	0	-
85-120	1	12.17	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
>120	1	10.77	1	32.09	0	0.00	0	0.00	0	0.00	0	0.00

Cohort 2	Length of last screen interval (months)											
Period 1	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	8	2.06	7	1.80	6	2.25	1	0.39	2	1.43	0	0.00
1 prev	2	1.52	0	0.00	0	0.00	0	0.00	0	0.00	0	-
10-36	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
37-60	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-
61-84	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
85-120	0	0.00	0	-	0	-	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-
Gap 2												
No prev	2	2.98	1	1.93	2	7.47	0	0.00	0	0.00	0	-
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
10-36	1	23.03	0	0.00	0	0.00	0	-	0	-	0	-
37-60	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
61-84	0	0.00	0	-	0	-	0	-	0	-	0	-
85-120	0	0.00	0	-	0	-	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-
Gap 3												
No prev	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
1 prev	0	0.00	0	-	0	-	0	-	0	-	0	-
10-36	0	-	0	-	0	-	0	-	0	-	0	-
37-60	0	-	0	-	0	-	0	-	0	-	0	-
61-84	0	-	0	-	0	-	0	-	0	-	0	-
85-120	0	-	0	-	0	-	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-

Cohort 2	Length of last screen interval (months)											
Period 2	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	32	2.14	39	1.94	28	1.64	38	1.50	32	1.24	6	0.77
1 prev	31	1.60	28	1.43	19	1.50	12	0.94	8	1.12	0	0.00
10-36	49	1.02	14	0.71	4	0.61	0	0.00	0	0.00	0	0.00
37-60	30	1.30	10	0.56	1	0.13	2	0.37	0	0.00	0	0.00
61-84	12	2.08	6	1.05	2	0.69	2	0.88	0	0.00	0	0.00
85-120	6	2.73	3	1.27	1	0.80	1	1.14	0	0.00	0	0.00
>120	2	3.02	1	1.71	0	0.00	0	0.00	0	0.00	0	-
Gap 2												
No prev	32	6.53	19	4.35	11	3.77	8	2.24	2	0.76	0	0.00
1 prev	18	2.79	13	2.73	7	2.61	1	0.40	0	0.00	1	5.28
10-36	14	0.99	6	0.99	3	1.52	0	0.00	0	0.00	0	0.00
37-60	10	1.50	4	0.89	3	1.47	0	0.00	0	0.00	0	0.00
61-84	3	1.73	1	0.81	0	0.00	2	3.10	0	0.00	0	-
85-120	4	5.72	1	2.18	0	0.00	0	0.00	0	0.00	0	0.00
>120	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
Gap 3												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-

10-36	1	5.76	0	0.00	0	0.00	0	0.00	0	-	0	-
37-60	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-
61-84	0	0.00	0	0.00	0	0.00	0	-	0	0.00	0	-
85-120	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
>120	0	0.00	0	0.00	0	-	0	-	0	-	0	-

Cohort 2	Length of last screen interval (months)											
Period 3	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	1	0.44	8	0.36
1 prev	1	1.83	3	2.81	1	0.76	3	1.06	2	0.33	7	0.57
10-36	29	0.47	10	0.28	5	0.33	6	0.47	2	0.18	0	0.00
37-60	10	0.35	14	0.40	9	0.41	5	0.22	1	0.06	0	0.00
61-84	3	0.40	5	0.41	3	0.29	5	0.41	3	0.27	1	0.20
85-120	3	0.83	3	0.47	3	0.54	9	1.12	1	0.13	0	0.00
>120	2	0.78	1	0.25	1	0.26	2	0.35	0	0.00	1	0.61
Gap 2												
No prev	0	0.00	1	23.76	0	0.00	0	0.00	1	1.13	0	0.00
1 prev	0	0.00	1	2.74	1	3.20	1	1.91	4	2.36	0	0.00
10-36	13	0.88	2	0.23	1	0.32	0	0.00	0	0.00	2	0.70
37-60	10	1.22	5	0.74	5	1.54	2	0.63	1	0.28	0	0.00
61-84	4	2.10	2	1.00	2	1.62	0	0.00	0	0.00	0	0.00
85-120	1	1.16	0	0.00	1	1.56	0	0.00	1	1.14	1	2.08
>120	1	1.65	1	2.20	0	0.00	2	4.23	0	0.00	0	0.00
Gap 3												
No prev	0	0.00	0	0.00	1	35.29	0	0.00	0	0.00	0	0.00
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
10-36	3	1.96	0	0.00	1	2.85	0	0.00	0	0.00	0	0.00
37-60	1	1.14	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
61-84	1	4.74	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
85-120	1	7.72	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
>120	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00

Appendix 2

Number of cases and prevalence rates per thousand individuals by:

- Cohort (1=born 1914-1918, 2= born 1929-1933)
- Period (1=pre-1963, 2= 1963-1975, 3=1976-1992)
- Degree of abnormality in Gap preceding last interval (1=no gap, 2=minor-<10 months or class 2 or 9, 3=major-class 3 or 4 or histology)
- Length of last screen interval,
- Combined length of two screens prior to last.
- Cells with no time-at-risk indicated with "-".

Cohort 1	Length of last screen interval (months)											
Period 1	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	9	2.30	4	1.57	3	2.51	1	1.14	0	0.00	0	0.00
1 prev	2	0.96	1	1.17	2	7.97	0	0.00	0	0.00	0	-
10-36	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
37-60	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
61-84	0	0.00	0	0.00	1	166.67	0	0.00	0	-	0	-
85-120	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
>120	0	0.00	0	-	0	-	0	-	0	-	0	-
Gap 2												
No prev	6	8.65	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
10-36	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-
37-60	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-
61-84	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
85-120	0	0.00	0	-	0	0.00	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-
Gap 3												
No prev	0	0.00	0	0.00	0	0.00	0	-	0	0.00	0	-
1 prev	0	0.00	0	-	0	-	0	-	0	-	0	-
10-36	0	-	0	-	0	-	0	-	0	-	0	-
37-60	0	-	0	0.00	0	-	0	-	0	-	0	-
61-84	0	-	0	-	0	-	0	-	0	-	0	-
85-120	0	-	0	-	0	-	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-

Cohort 1	Length of last screen interval (months)											
Period 2	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	11	1.88	12	2.12	11	3.01	14	3.46	7	2.38	6	8.72
1 prev	10	1.13	12	2.05	3	1.11	5	2.33	6	6.29	2	14.29
10-36	13	0.45	5	0.72	1	0.62	1	1.29	0	0.00	0	0.00
37-60	13	1.09	2	0.32	1	0.53	2	1.92	0	0.00	0	0.00

61-84	3	1.06	2	1.00	1	1.32	0	0.00	1	6.90	0	0.00
85-120	2	1.54	1	1.21	2	5.95	2	10.81	0	0.00	0	0.00
>120	1	2.15	1	3.79	1	9.90	0	0.00	0	0.00	0	-
Gap 2												
No prev	5	3.21	2	2.20	3	7.06	3	6.55	0	0.00	1	14.71
1 prev	2	0.84	1	0.86	0	0.00	3	8.72	2	13.24	0	0.00
10-36	6	0.80	4	2.01	1	2.62	0	0.00	1	14.09	0	0.00
37-60	5	1.61	2	1.51	0	0.00	0	0.00	0	0.00	0	0.00
61-84	3	3.89	1	2.38	0	0.00	0	0.00	0	0.00	0	0.00
85-120	0	0.00	1	5.62	0	0.00	0	0.00	0	0.00	0	-
>120	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
Gap 3												
No prev	2	40.82	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
1 prev	1	21.28	1	62.50	0	0.00	0	0.00	0	-	0	-
10-36	1	9.26	0	0.00	0	0.00	0	0.00	0	0.00	0	-
37-60	1	17.54	1	100.00	0	0.00	0	0.00	0	0.00	0	-
61-84	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
85-120	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
>120	0	0.00	0	0.00	0	-	0	-	0	-	0	-

Cohort 1	Length of last screen interval (months)											
Period 3	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	5	20.66	6	5.53
1 prev	0	0.00	1	3.42	4	15.94	0	0.00	0	0.00	4	5.47
10-36	5	0.25	5	0.70	0	0.00	1	0.60	1	0.90	2	3.82
37-60	5	0.57	3	0.45	2	0.63	3	1.10	2	1.13	0	0.00
61-84	4	1.60	7	2.77	2	1.26	2	1.26	3	2.75	1	2.97
85-120	1	0.69	1	0.69	2	2.02	1	0.94	2	2.46	1	4.33
>120	0	0.00	4	3.42	0	0.00	0	0.00	2	3.02	0	0.00
Gap 2												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	6.41
1 prev	0	0.00	0	0.00	0	0.00	1	13.51	0	0.00	1	8.62
10-36	3	0.72	4	2.43	2	4.26	0	0.00	0	0.00	0	0.00
37-60	1	0.45	1	0.81	1	2.09	1	2.72	2	7.69	1	11.24
61-84	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
85-120	0	0.00	0	0.00	1	7.75	0	0.00	1	11.77	0	0.00
>120	1	3.53	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Gap 3												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
10-36	1	2.53	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
37-60	1	3.80	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
61-84	1	12.19	0	0.00	0	0.00	0	0.00	0	0.00	0	-
85-120	1	15.87	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
>120	1	17.24	1	62.50	0	0.00	0	0.00	0	0.00	0	0.00

Cohort 2	Length of last screen interval (months)											
Period 1	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	8	2.66	7	3.75	6	7.07	1	1.71	2	9.90	0	0.00
1 prev	2	1.93	0	0.00	0	0.00	0	0.00	0	0.00	0	-
10-36	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
37-60	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-
61-84	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
85-120	0	0.00	0	-	0	-	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-
Gap 2												
No prev	2	3.98	1	4.10	2	24.10	0	0.00	0	0.00	0	-
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
10-36	1	30.30	0	0.00	0	0.00	0	-	0	-	0	-
37-60	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
61-84	0	0.00	0	-	0	-	0	-	0	-	0	-
85-120	0	0.00	0	-	0	-	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-
Gap 3												
No prev	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
1 prev	0	0.00	0	-	0	-	0	-	0	-	0	-
10-36	0	-	0	-	0	-	0	-	0	-	0	-
37-60	0	-	0	-	0	-	0	-	0	-	0	-
61-84	0	-	0	-	0	-	0	-	0	-	0	-
85-120	0	-	0	-	0	-	0	-	0	-	0	-
>120	0	-	0	-	0	-	0	-	0	-	0	-

Cohort 2	Length of last screen interval (months)											
Period 2	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	32	2.89	39	4.23	28	5.21	38	6.96	32	9.19	6	9.59
1 prev	31	2.16	28	3.12	19	4.84	12	4.32	8	8.06	0	0.00
10-36	49	1.34	14	1.52	4	1.96	0	0.00	0	0.00	0	0.00
37-60	30	1.77	10	1.22	1	0.42	2	1.66	0	0.00	0	0.00
61-84	12	2.88	6	2.30	2	2.21	2	3.99	0	0.00	0	0.00
85-120	6	3.72	3	2.83	1	2.53	1	5.24	0	0.00	0	0.00
>120	2	4.08	1	3.73	0	0.00	0	0.00	0	0.00	0	-
Gap 2												
No prev	32	9.62	19	9.87	11	12.11	8	10.28	2	5.68	0	0.00
1 prev	18	4.12	13	6.12	7	8.55	1	1.83	0	0.00	1	62.50
10-36	14	1.42	6	2.20	3	4.93	0	0.00	0	0.00	0	0.00
37-60	10	2.20	4	1.97	3	4.80	0	0.00	0	0.00	0	0.00
61-84	3	2.60	1	1.81	0	0.00	2	13.89	0	0.00	0	-
85-120	4	8.62	1	4.74	0	0.00	0	0.00	0	0.00	0	0.00
>120	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
Gap 3												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	-
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-

10-36	1	7.41	0	0.00	0	0.00	0	0.00	0	-	0	-
37-60	0	0.00	0	0.00	0	0.00	0	0.00	0	-	0	-
61-84	0	0.00	0	0.00	0	0.00	0	-	0	0.00	0	-
85-120	0	0.00	0	0.00	0	0.00	0	-	0	-	0	-
>120	0	0.00	0	0.00	0	-	0	-	0	-	0	-

Cohort 2	Length of last screen interval (months)											
Period 3	10-18		19-30		31-42		43-54		55-120		>120	
	n	rate	n	rate	n	rate	n	rate	n	rate	n	rate
Gap 1												
No prev	0	0.00	0	0.00	0	0.00	0	0.00	1	3.77	8	6.32
1 prev	1	2.76	3	6.55	1	2.46	3	5.09	2	2.62	7	9.35
10-36	29	0.60	10	0.60	5	1.02	6	2.11	2	1.34	0	0.00
37-60	10	0.46	14	0.85	9	1.30	5	0.99	1	0.42	0	0.00
61-84	3	0.54	5	0.89	3	0.92	5	1.83	3	1.96	1	2.70
85-120	3	1.11	3	1.03	3	1.70	9	5.09	1	0.93	0	0.00
>120	2	1.03	1	0.54	1	0.83	2	1.59	0	0.00	1	7.46
Gap 2												
No prev	0	0.00	1	47.62	0	0.00	0	0.00	1	8.48	0	0.00
1 prev	0	0.00	1	6.49	1	10.64	1	9.35	4	18.69	0	0.00
10-36	13	1.26	2	0.52	1	1.01	0	0.00	0	0.00	2	10.20
37-60	10	1.73	5	1.63	5	4.91	2	2.83	1	2.11	0	0.00
61-84	4	2.98	2	2.24	2	5.11	0	0.00	0	0.00	0	0.00
85-120	1	1.62	0	0.00	1	5.18	0	0.00	1	8.40	1	27.78
>120	1	2.32	1	4.72	0	0.00	2	19.61	0	0.00	0	0.00
Gap 3												
No prev	0	0.00	0	0.00	1	200.00	0	0.00	0	0.00	0	0.00
1 prev	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
10-36	3	2.86	0	0.00	1	9.01	0	0.00	0	0.00	0	0.00
37-60	1	1.62	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
61-84	1	6.54	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
85-120	1	12.35	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
>120	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00