NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

A CASE-CONTROL STUDY OF RISK FACTORS FOR ECTOPIC PREGNANCY USING ROUTINELY COLLECTED ADMINISTRATIVE DATA

by

JAMES CHARLES HOCKIN

Thesis submitted to
the School of Graduate Studies and Research
in partial fulfilment of the requirements for the
M.Sc. degree in Epidemiology

University of Ottawa

October 10, 1990

ISBN 0-315-68106-3

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

Abstract

Data collected for the administration of health care programs are frequently used in epidemiological research. The most useful administrative data for this purpose are those which provide information on individual claims payments or hospitalizations. In two Canadian provinces, the agencies which manage these data have developed procedures for linking claims, hospital and other records for research purposes based on unique subscriber numbers. The ability to link hospital, mortality, and claims data sets across time makes it possible to study individuals without resorting to time-consuming and expensive interviews or chart review. This form of record linkage has been used in a number of published studies, most of which use a cohort design because data are readily available for large populations.

The case-control design has had limited use in such studies because of the potential bias arising from poor ascertainment of exposures and/or confounders. The availability and low cost of obtaining data for large populations make a case-control design appealing when one wishes to consider multiple exposures which are represented in the data. The hospital separations database for most provinces is readily accessible to researchers in a form which does not permit identification of individuals, yet which permits internal record linkage of hospital separations for an individual. Records available from the Hospital Medical Records Institute (HMRI) in Canada contain date of birth, encrypted health care identifier, sex, date of separation, diagnostic and procedure codes and other variables related to hospital stay. The potential for

using the case-control design with only hospital separation records was explored in a study of ectopic pregnancy.

The Saskatchewan Hospital Services Commission hospital morbidity database was used. The database dates back to 1970 and records all hospital separations for a fairly stable population of about one million. Cases were women hospitalized for ectopic pregnancy in the years 1984 through 1986. Controls were women hospitalized for any other pregnancy condition, frequency matched on 5-year age group and time of hospital separation. Hospital records were internally linked by the provincial hospital services plan, providing a longitudinal record beginning in 1970 or the time of arrival in the province for each individual. Secondary diagnosis and procedure codes within the data validated the majority of incident episodes of ectopic pregnancy in the study period.

In a case-control study using pregnant controls, the odds ratio is an appropriate estimator of relative risk, conditional on conception. Among the potential risk factors (exposures) represented in the data were tubal surgery, prior ectopic pregnancy, pelvic inflammatory disease, and infertility. Of these, the first two are completely ascertained during the time a woman was resident in Saskatchewan. Under representation of exposures should be non-differential, leading to odds ratios which are biased towards the null. Multiple logistic regression analysis yielded estimates of the odds ratios for acute PID (2.3, 95% confidence interval [CI] 1.1-4.9), chronic PID or pelvic adhesions (2.7, CI 0.8-9.2), prior ectopic pregnancy (8.9 CI 3.0-26.3), tubal sterilization (9.4, CI 1.3-69), and tubal repair (12.4, CI 1.8-87). Of particular interest was the association between ectopic pregnancy and tubal surgery performed in the hope of maintaining or restoring normal tubal function. The attributable risk proportion of ectopic pregnancy due to such surgery was 8.1%. This study design can be used to follow trends in surgical causes of ectopic pregnancy in Canada.

Acknowledgements

Conclusions drawn from the study reported here are those of the author and are not necessarily those of the **Saskatchewan Hospital Services Plan.**

I am indebted to the staff of the Systems and Data Processing division, Saskatchewan Hospital Services Plan, for their cooperation and suggestions made during negotiations for acquisition of the data utilized in this study.

The assistance of the Department of National Health and Welfare, in the form of educational leave to pursue graduate studies at the University of Ottawa, is gratefully acknowledged.

The guidance of my supervisor, S. Raman, and of R.A. Spasoff and P. Stewart, of the University of Ottawa, Department of Epidemiology and Community Medicine, is appreciated.

My studies of the occurrence of ectopic pregnancy in Canada were inspired by work done by A.G. Jessamine and the late Marion Todd of the STD Division, Bureau of Communicable Disease Epidemiology, at the Laboratory Centre for Disease Control.

Table of Contents

List of Tables

List of Figures

Introduction

Rationale for studying ectopic pregnancy trends

An **ectopic pregnancy** is defined as the implantation of a fertilized ovum anywhere outside of the uterine cavity. Most ectopic implantations occur in the fallopian tube, which is unable to sustain growth of the embryo beyond a few weeks. The growing trophoblastic tissue is invasive and demands a rich maternal vascular supply. Wnen the embryo grows beyond a certain size, it can no longer be supported in the abnormal location and invariably dies. Frequently, the fallopian tube will rupture and life-threatening bleeding may follow[1]. Ruptured ectopic pregnancy has long been a leading cause of maternal mortality in some populations[2].

Since 1970, there has been a global trend to increasing incidence of ectopic pregnancy[3]. An understanding of the factors leading to the increasing trend would assist in the planning and prioritizing of strategies to prevent ectopic pregnancy, yet there has been relatively little effort to determine which of the known risk factors may be responsible for the trend. It is possible that new diagnostic technology and education of health professionals have led to an increase in diagnosis to the point where previously unrecognized, spontaneously resolving, ectopic pregnancies are now diagnosed and treated.

The absolute or relative risk of ectopic pregnancy given a known risk factor is of prognostic value to clinicians and to patients who desire to become pregnant. In contrast, population attributable risk is more relevant to prevention

of the escalating morbidity and cost of health services related to ectopic pregnancy. In recent years, a number of population-based studies have estimated the relative risk of ectopic pregnancy following salpingitis and other exposures. None has been able to take account of changes over time in order to explain satisfactorily the trend over the past 20 years. None has attempted to estimate population attributable risk.

Canadian trends

Trends in hospital separations of ectopic pregnancy in Canada have been published for 1971 through 1983/84[4,5]. Data are now available for fiscal year 1985/86 (Health Care Section, Statistics Canada, unpublished data). Separations for ectopic pregnancy among women age 15-44 in Canada increased from 5.7 per 1000 measurable conceptions in 1971 to 13.9 per 1000 in 1985/86 (Figure 1). **Measurable conceptions** include live births, stillbirths, induced abortions and ectopic pregnancies. There are similar trends in reported gonorrhoea and pelvic inflammatory disease and some authors have explained the trend in ectopic pregnancies as a direct consequence of the epidemic of salpingitis due to gonorrhoea and chlamydia. Separation rates standardized to the 1971 age distribution of reported pregnancies differ little from crude rates, accounting for only 14% of the increase over 15 years (Figure 2). There is a strong cohort effect with higher rates at all ages for younger cohorts (Figure 3).

Use of administrative data to study etiology

Administrative databases make available data on large populations over long periods of time. They provide a potentially rich harvest of raw data on which to build a conventional epidemiological study. An example of such a database is the Saskatchewan Hospital Services Plan (SHSP) record of hospital separations from 1970 onward. Until recently, the vast majority of known,

symptomatic, ectopic pregnancies required a hospital stay for definitive (surgical) treatment[2]. The hospital discharge database for a province should, therefore, contain a record of virtually all symptomatic, non-fatal cases of ectopic pregnancy in a defined population.

The thesis proposes to examine three questions which relate to the use of administrative data for studying ectopic pregnancy:

(1) **Can hospital morbidity data and case-control methodology be used to estimate the relative risk of ectopic pregnancy (EP) due to pelvic inflammatory disease (PID), or other risk factors, treated in hospital?**

(2) **Can etiologic fraction (population attributable risk fraction) be estimated from such a case-control study, with or without additional data?**

(3) **Is there a detectable trend in the estimated etiologic fractions for EP occurring in the period 1984 to 1987?**

While some risk factors for ectopic pregnancy are not completely ascertained in hospital separation data, the low cost of obtaining large amounts of data permits the design of a study with adequate power. The thesis will discuss the limitations and potential use of such data in estimating etiologic fractions for ectopic pregnancy risk factors. If the data and methods are valid, then the study can be a useful surveillance tool which, when repeated, will lead to a better understanding of the trend in ectopic pregnancy. Finally, the thesis will consider other research hypotheses which may be examined with similar methodology.

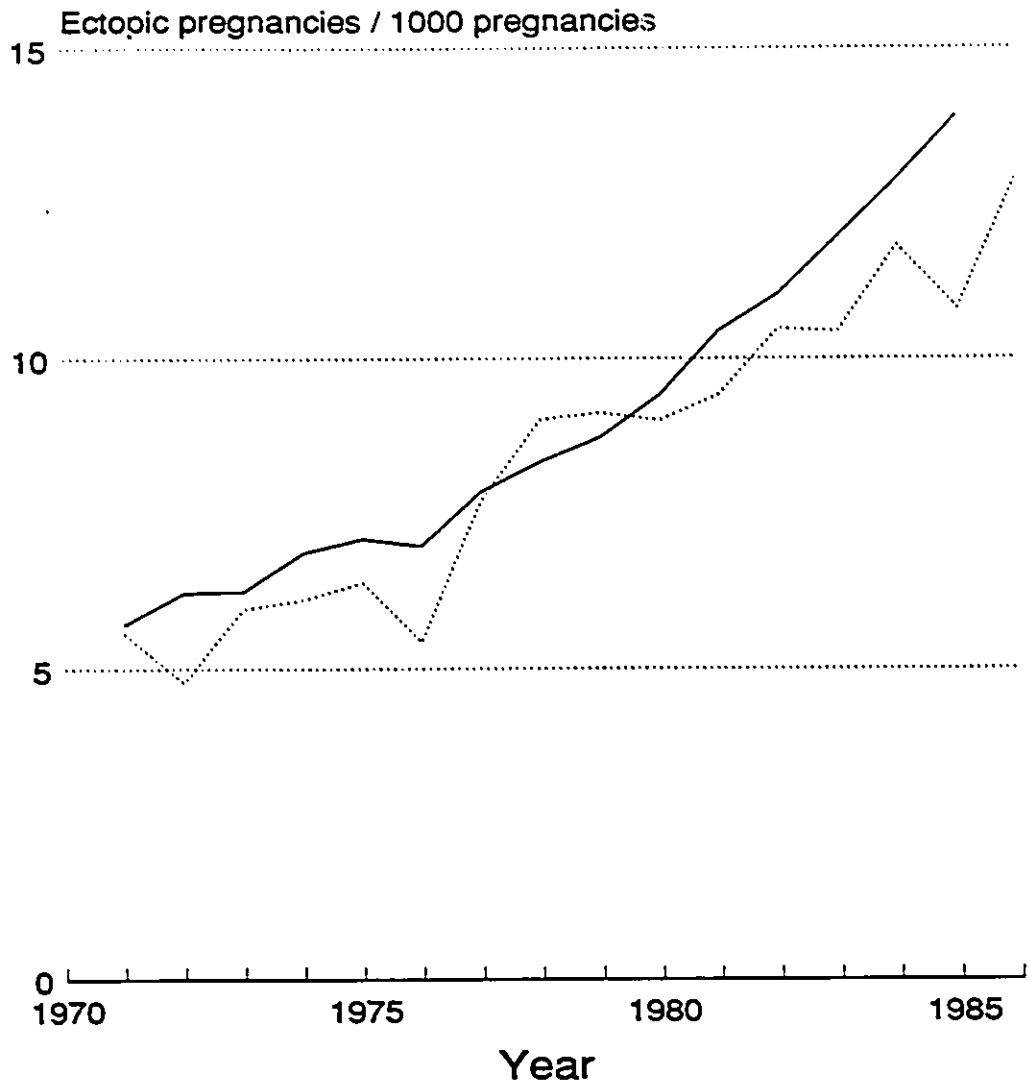

Figure 1. Ectopic pregnancy, Canada and Saskatchewan, 1971-1986

Hospital separations for ectopic pregnancy per 1000 reported pregnancies (live births, stillbirths, induced abortions and ectopic pregnancies); solid line is Canada; dotted line is Saskatchewan. Data for Saskatchewan, 1984-1986 are from SHSP data; remainder of data are from Statistics Canada.

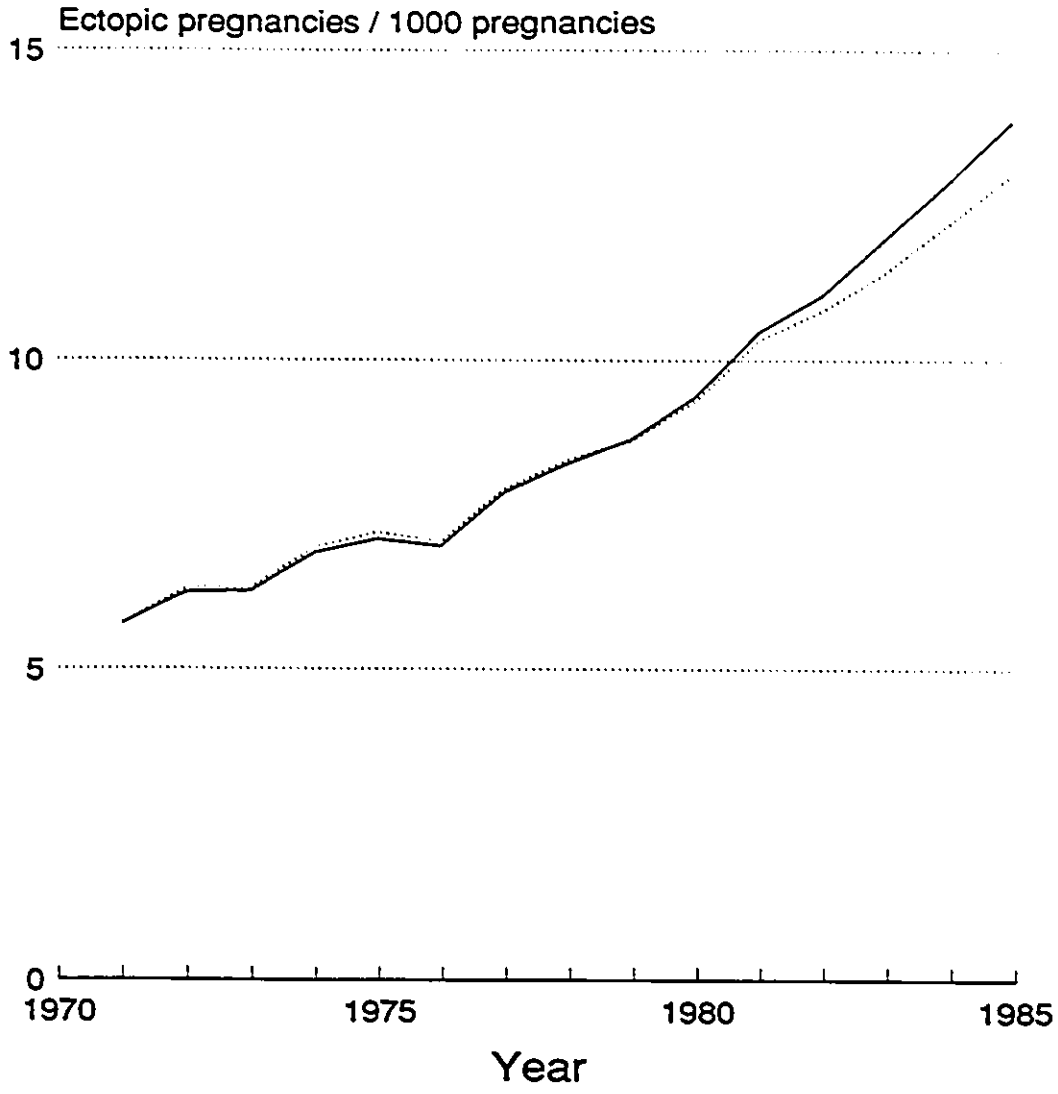

Figure 2. Ectopic pregnancy, Canada, 1971-1985

Hospital separations for ectopic pregnancy per 1000 reported pregnancies (live births, stillbirths, induced abortions and ectopic pregnancies); solid line is crude rate; dotted line is rate directly standardized to the 1971 age distribution of reported pregnancies.

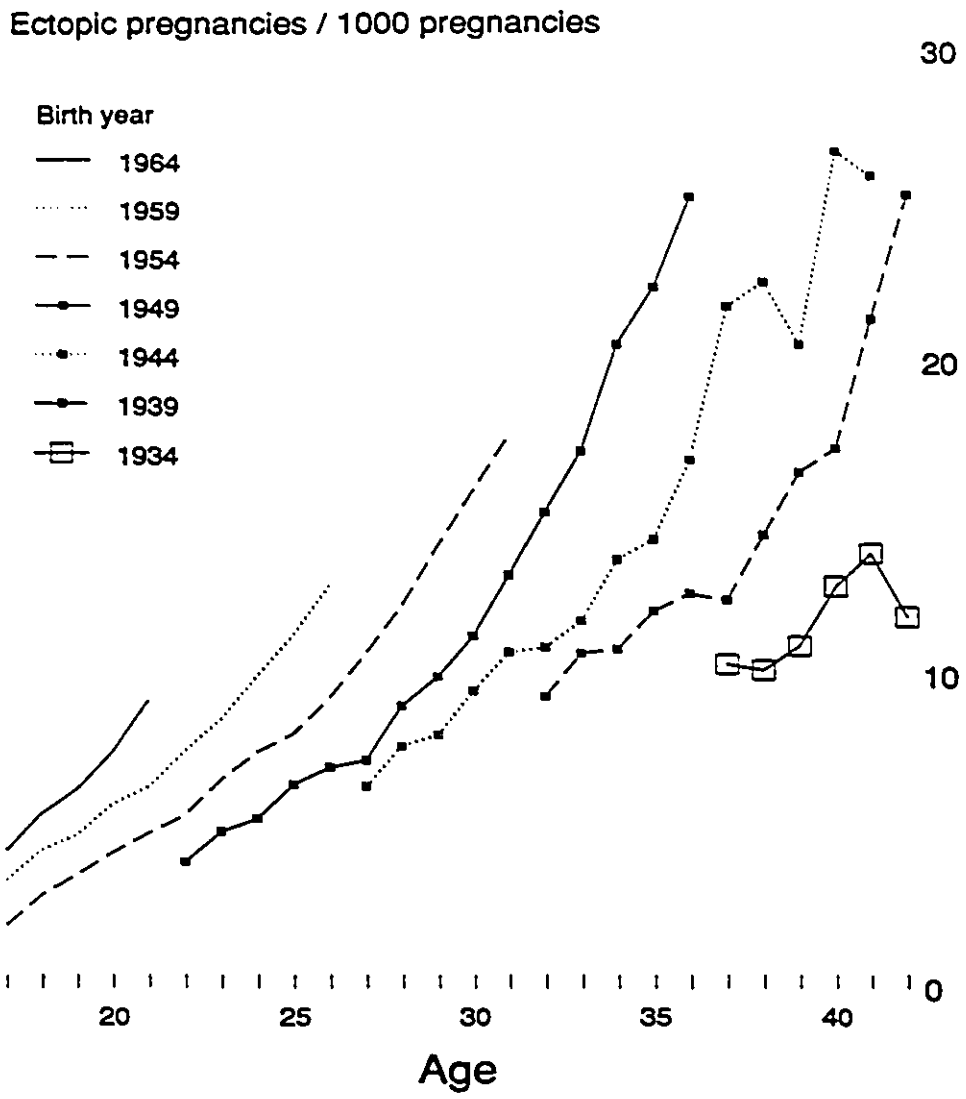

Figure 3. Ectopic pregnancy for birth cohorts, Canada, 1971-1985

Hospital separations for ectopic pregnancy per 1000 reported pregnancies (live births, stillbirths, induced abortions and ectopic pregnancies) for selected birth cohorts; rates are interpolated from age-specific rates for 5-year age groups.

CHAPTER I

Literature Review

Ectopic Pregnancy trends and risk factor studies

Because the primary interest of this thesis is methodological, the theme of the ectopic pregnancy literature review relates primarily to data sources (administrative & vital statistical data, chart review, patient interview, pathological, clinical follow up), and method of case/control identification. I have reviewed the literature in terms of study type, case and referent populations, and factors studied.

In a 1982 paper, Chavkin[6] reviewed case series and case-control studies which reported associations with proposed risk factors for EP, in an attempt to explain the increasing incidence. She concluded that "many of [the studies] suffer from methodological flaws", and made a plea that "future research [develop] study designs to yield results that could have direct impact on clinical practice and preventive measures." An excellent and comprehensive review by Chow in 1987 summarized the studies of the etiology of EP[7].

Descriptive studies

Descriptive epidemiology studies of EP have been based on case series or surveillance data. Prior to 1980, studies of ectopic pregnancy epidemiology were almost exclusively based on case series in gynaecological practices or hospitals[3,7]. In the 1980s, there were at least 52 reports of case series

worldwide, including 12 which describe follow up on pregnancy outcomes*. Experience with cases of ruptured ectopic pregnancy led to publication by a number of authors of clinical reports of the "classical" presentation of EP which may be found in many textbooks[2]. The frequency of the classical signs and symptoms described in case series have changed since the introduction of new diagnostic tools which permit definitive diagnosis of EP prior to rupture[8,9,10].

Barnes et al reviewed the incidence of EP, as reported prior to 1980, in 1983 and noted that different denominators had been used to estimate the incidence of EP in various surveillance reports[3]. In many cases, including the earliest of their references[11], EP incidence was estimated by the ratio of EP to total births (live plus stillborn). Other denominators reported included female population of childbearing age, live births, and diagnosed conceptions. Barnes et al recommended that the best denominator to adopt for comparison purposes was the female population age 15-44. Comparisons could only be made, however, if incidence were standardized to some population. However, Barnes et al did point out a significant inverse correlation between births and induced abortions among the 50 states (r=-0.49). This underlines the importance of including all pregnancy outcomes in the denominator. Apparent trends in rates which use only live births as a denominator could otherwise be an artifact of a changing incidence in elective termination of pregnancy.

Since 1980, there have been at least 21 reports which include time series data, plus a series of surveillance reports from the Centers for Disease Control[12,13,14,15,16,17]. These showed that the incidence was increasing by 4.8% to 13.9% per year. This is similar to the Canadian experience in the decade ending 1980[4]. Surveillance summaries from the CDC and LCDC have consistently used both population and "reported pregnancies" (live births,

---

*based on a Medline search of articles indexed on "ectopic pregnancy" and "epidemiology". Many, but not all, reports are cited below.

stillbirths, induced abortions and ectopic pregnancies) as denominators. The latter denominator takes into account changes in fertility by age, contraception, and abortion, all of which would bias ratios or proportions using other denominators. Other reports of pregnancy outcomes for geographic populations have shown an increasing incidence of EP through the 1970s and 1980s in the United States[18,19,20,21,22], Finland[23,24,25], Sweden[26], Denmark[27], Iceland[28], Scotland[29], New Zealand[30,31], Australia[32], Tasmania[33], Italy[34], and Poland[35].

There have been a number of reports which include time series data for a single hospital. Gorodeski & Bahary found a jump in 1975 in the number of EP at one Tel Aviv hospital and almost all of the increase was due to increasing numbers of cases with unruptured tubal pregnancy[36]. Zhang, at a Chinese hospital, found that the proportion of EP/1000 conceptions jumped in 1983-84, when the number of IUD insertions increased 5-fold[37]. Almost all of the increase in EP was attributed to IUD use at the time of conception. Each study was limited to the experience of a single hospital which may not be representative of the population.

The only reported exception to an increasing incidence of EP in the 1970s was in Enugu, Nigeria, where Egwuatu and Ozumba found that the ratio of EP:live births fell from 1:190 in 1978 to 1:480 in 1981[38]. However, for reasons already mentioned, total live births is not an appropriate denominator and the trend is based only on a single teaching hospital, so may not be representative of the population in Enugu, or Nigeria as a whole.

Routinely gathered hospital morbidity data may underestimate the incidence of EP. Several authors have pointed out that a tubal pregnancy may abort and resolve without intervention[2,18]. Prior to the introduction of modern diagnostic tools (ultrasonography, laparoscopy, ß-HCG assay), such EP may have gone undetected. Improved diagnosis may increase the probability of detection of these EP and lead to surgery, hospital stays, and a hospital

separation record. If so, then these technological changes could account for part of the increasing trend. The implication for surveillance is that an undetermined proportion of EP are of the potentially resolvable type. Hospital morbidity data thus provide estimates of the incidence of recognized EP. If the undetected ectopic pregnancies are considered clinically important, then failure to include them as outcomes will reduce the power of cohort studies.

Case series

Many reports of case series include data on risk factors leading to EP. The data for these studies are derived from chart and pathological review and the implication is that a case may be attributed to one or more known factors. This may seem reasonable to the clinician who must consider the prognosis regarding future tubal function, but it cannot be uses directly to estimate attributable risk, which is of public health interest. There is also the potential for selection bias in case series. For example, in a small series accumulated over 30 years in their Philadelphia practice, Corson and Batzer report on 103 cases[39]. Of these, only 4 (4%) had an IUD in situ. Weinstein et al reported another small series gathered over a 10-year period in Tucson, in which 15 (9.7%) had an IUD in situ[40]. Over a similar period in Finland, however, Tuomivaara et al found that 158 of 458 (34%) of their cases had an IUD in situ[41]. Similar variation can be found for other risk factors (PID, infertility, tubal ligation, etc) depending on the population-time studied and the selection biasses which operate in individual practices or hospitals. None of these studies provides estimates of absolute, attributable, or relative risk which could be of prognostic value. All suffer from the lack of a comparison group and well defined denominators.

Case-control studies

The major difficulty in the design of case-control studies of EP is the selection of controls appropriate to the conditional risks of interest[42]. There are two situations to consider. In the first, one wishes to estimate relative risk of EP given some exposure, conditional only on being sexually active. This is appropriate for methods of contraception. Exposure (contraceptive use) may not necessarily imply sexual activity, but this is likely. Controls must be selected from a population of sexually active women, or analysis must control for sexual activity by exclusion. The IUD and oral contraceptives have been studied and reviewed extensively[6,43,44]. Seven of the case-control studies reviewed by Chavkin[6] investigated IUD use. In three of these, the domain of study was restricted to IUD users and consideration was given to those factors, such as salpingitis, which increased the risk of EP. In one other, correlations of other risk factors with IUD use were studied in women having ectopic pregnancy with or without an IUD in situ.

The second situation is a study of the relative risk of EP conditional on conception. Risk estimates are for the outcome of ectopic implantation of a fertilized ovum. Controls must be selected from a population of women who have conceived. Many factors (contraception, PID, infertility) are inversely associated with the conditioning factor, conception. A pregnant control group will have proportionally fewer women with these risk factors. Therefore, risk estimates will be higher than in the first situation. Such estimates may be more appropriate for prognosis of women trying to conceive in the presence of risk factors, because they will indicate the need for management of early pregnancy.

Either method of control selection is appropriate for estimation of attributable risk. Pregnant controls are required to estimate the proportion of actual cases (etiologic fraction) attributable to a factor; population controls will permit estimation of the population attributable risk (or preventive fraction).

A number of recent case-control studies will be considered in more detail. Table 23, page 79, summarizes results from the major studies, which are also included in the review by Chow[7].

Seattle study[45,46,47,48,49,50]. There are multiple papers based on this study. One reason for the many papers is that in estimating relative risk for a given factor (smoking), the authors have adjused for a different set of potential confounders, and have taken care **not** to adjust for factors which may also be in the same causal pathway.

Cases were married women treated for EP at one of five King County, WA hospitals. Cases were identified by medical records departments based on discharge diagnosis. Case inclusion criteria depended on the factors being studied, but overall, of 312 cases identified, 192 (61%) participated. The majority of non-participants could not be located; few refused. Controls had been selected in an earlier study of infertility, for which the interview was identical. Pregnant population controls were identified from vital statistical records of live birth and were either married or living in a stable relationship. This process excluded women having other pregnancy outcomes. Controls had been matched on age and other variables to cases in the original study, but no matching was employed in the study of EP. Of 607 controls identified, 459 (76%) participated. There was potential selection bias, since cases were drawn from only some of the hospitals serving the population from which controls were selected. Non-married controls were less likely to have been selected. Non-respondance would bias the results to null if it occurred non-differentially. Recall or response bias may affect some of their variables, for example, prior induced abortion, smoking, IUD use, history of STD or PID.

Depending on the factor of interest, cases and controls were restricted. For example, in the estimation of relative risk due to prior infertility, excluded were those who were using a contraceptive at the time of conception and those

whose partners had infertility. They justify control selection based on live births by restricting the study to married couples (and those in a stable relationship). The ratio of live births to induced abortions among married women in Seattle was 1:0.11; among single women it was 1:2.5. This was, however, post hoc planning as the control group belonged to another study.

Mayo Clinic study[51]. Cases were identified from Mayo Clinic records. All were white women who had been resident in Rochester, Minnesota at the time of diagnosis. Excluded were women having a prior EP. Cases occurred from 1935 through 1982. Two controls delivering a live infant were matched on date of conception, residency, race, prior ectopic, 5-year age stratum for each case. Major drawbacks to this study are a relatively small sample size, exclusion of women having induced abortion from the control group, and the long time period covered.

The Women's Health Study[52,53]. Data for this study were derived from a multicenter case-control study of the relationship between IUD use and a number of obstetric and gynaecological disorders occurring over a 24-month period in 1976-78. EP was one of the conditions considered at the design stage. Cases were identified by prospective surveillance of admission records and other hospital records for possible EP. Excluded were women who themselves or whose partners had been sterilized, and women for whom a medical diagnosis would influence contraceptive method. Controls were sexually active women, not pregnant or post-partum, recruited from medical and surgical services of the same hospitals. Exclusion criteria were the same as for cases. This was a classic study estimating relative risk for EP of IUD use, conditioned on sexual activity, but not on pregnancy.

WHO study[44]. This international study included 1108 cases and 2216 controls from participating clinics in developed and third-world countries. Each case was matched to a pregnant and a non-pregnant control.

Zhang[37] compared current contraceptive histories of 259 cases to two non-pregnant control groups, one from non-gynaecological hospital admissions, the other from the same neighbourhood, matched on age and parity. Ninety-nine cases had an IUD in situ; most were steel ring IUD's common in China. Neither control group meets the criteria set out above for the study of EP, and the analysis did not control for sexual activity.

Thorburn et al[54,55] used stepwise logistic regression to estimate the minimal model useful in predicting EP risk in a case-control study done in Gothenburg, Sweden. Because 25% of pregnancies were terminated in that city, they had two unmatched control groups, one composed of pregnant women desiring to deliver and the other, women having induced abortions. Significant univariate odds ratios were found for prior EP, prior D&C, history of salpingitis, prior abdominal surgery, infertility for more than 1 year, and IUD in situ. Four factors entered their final model: EP, IUD, infertility, abdominal surgery. Using population data on the distribution of pregnancy outcomes and Bayes' theorem[56], they modelled the predicted probability of an ectopic pregnancy, given various factors, with a view to prognosis. Predicted probabilities ranges from 1% to greater than 25%.

Handler et al[57] used data from an existing perinatal registry at the University of Illinois to study smoking and EP. The registry is based on ongoing chart abstraction. Controls were all women delivered of a single live-born child. Pregnant controls are appropriate because smoking is hypothesized to affect tubal transport and fertility, so relative risk estimation should be conditional on conception. They found an odds ratio of 2.5 for current cigarette smokers as well as a gradient for amount smoked.

All of these case-control studies utilized patient interviews or chart abstraction to gather data. Only the Thorburn study was based on a geographically defined population.

Cohort studies

The majority of cohort studies of ectopic pregnancy have been follow up studies of the pregnancy outcomes of uncontrolled series of cases treated by one or more surgical techniques. These studies may relate risk factors to outcome but there is confounding present since identified risk factors for EP may have dictated treatment.

Weström et al[58] studied the population of Lund, Sweden where all obstetrical and gynaecological services are provided by the university hospital. From 1960 to 1979 they treated 249 cases of EP and managed 30,822 intrauterine pregnancies. In a follow-up study of women having laparoscopically proven salpingitis they were able to estimate the risk of EP following salpingitis. This risk increased with increasing numbers of episodes of acute salpingitis. They also determined that there was no increased risk of EP during IUD use (relative to non-contraceptors) in the absence of proven salpingitis.

Ecological analyses

Beral[59] estimated rates of EP in England and Wales as EP/population and EP/(live births plus induced abortions) for the period 1964-72. In an attempt to understand the reasons for the sharp increase since 1970, she compared the percentage increase in rates, using a population denominator, for 1971/72 over 1968/70, in 14 regions, to the change in numbers of hospital diagnoses of salpingitis/oophoritis, the estimated proportion of post-partum women using an IUD and the incidence of induced abortion. In a rather weak ecological study, she found a correlation of +0.60 between percent increases in EP and use of IUD. The latter was estimated from a small survey done in 1973.

Etiology of ectopic pregnancy

Those factors supported in the literature, based on unbiased studies or pathology review include contraceptives (IUD, progestagen only mini-pill), infection, congenital anomalies, (including DES-induced fallopian tube anomalies), surgery, (including tubal ligation and tubal repair), and smoking. The reader is referred to the review by Chow for a full consideration of risks. The increased risk due to intra-uterine devices is conditional on conception. That is, among women using an IUD, there is an increased risk of ectopic implantation of a fertilized ovum, but the risk of ectopic pregnancy is lower than for non-contracepting, sexually active women.

Etiological fractions have not been reported in studies of etiology, but pathology studies provide upper limits of etiologic fractions for some factors in the populations studied.

Table 1 summarizes the study types and data sources encountered in this review. It may be noted that etiological studies using only administrative data have not been attempted.

Table 1. Types and data sources for studies of ectopic pregnancy 1980-89

	Study type			
Data source	Descriptive	Ecologic	Case-control	Cohort
Chart/path	Yes		Yes	Yes
Interview	Yes		Yes	Yes
Administrative only	Yes	Yes		
Mixed	Yes	Yes	Yes	Yes

Use of administrative data for research

For the purpose of this review, **population-based health data** can be considered to fall into one of three categories. **Administrative data** are financial in nature and are derived from administrative records maintained on individuals and which are used by government for the administration of programs directed at the population[60]. **Vital statistics** are legal in nature and are derived from data collected specifically for the purpose of recording and tabulating the occurrence of life events (births, deaths, marriages, divorce). **Epidemiological research data** are collected for the express purpose of describing or understanding relationships between health, disease and attributes of populations or individuals. All three types of data are used to describe the health status of whole populations.

Hospital separation is the event which is used to record the occurrence of a stay in hospital. Separation means discharge from hospital, transfer to another institution or death in hospital. It does not occur, by definition, for emergency room or hospital clinic visits or when surgical procedures are done without an overnight stay in hospital. In Canada, records of all acute care, convalescent, and chronic care hospital separations are abstracted, primarily to provide economic data to hospital administrators and provincial departments of health[61]. This data collection is universal in the ten provinces. Publicly funded hospitals must provide this information as a requisite to receive funding. The majority of hospital records for Canada are processed by the Hospital Medical Records Institute (HMRI), a non-profit organization. HMRI processing includes internal validity checks of the data and hospitals have an opportunity to correct coding errors. HMRI now collects records for almost all hospitals in all provinces except Nova Scotia, Quebec and Manitoba. Each province ultimately receives complete hospital records, regardless of the abstraction process.

Provincial records for a complete fiscal year are assembled and forwarded to Statistics Canada. The data are published annually, three to four years following the close of the fiscal year, as counts of separations and lengths of stay, by diagnosis, age, sex and province[61]. Tables for each calendar year are available by 3-digit ICDA-8 code from 1970 to 1978. Since then, they have been published by fiscal year using the Canadian Diagnostic List, an abridged list which groups 3-digit ICD-9 codes into 211 categories.

Tabulations of hospital morbidity in Canada have been used to gauge the relative occurrence of various conditions which require hospitalization[62,63]. In many instances, these data are not adequate to estimate incidence, because they refer to hospital episodes, not incident cases, of illness. For most chronic diseases, for which multiple hospitalizations occur for an individual over time, it is not reasonable to assume that each hospital separation corresponds to a unique episode of illness. On the other hand, disease episodes and hospitalizations correspond well for conditions representing the outcome of pregnancy, whether the outcome is delivery or abortive[4,18]. This is also true of many injuries which are resolved with a single stay in hospital[64]. Even for these conditions, however, hospital morbidity data may include some duplication because of transfers between hospitals. The coding of morbidity data do provide for unduplicating transfers, but this has been done infrequently[65].

The source databases for vital statistical and morbidity summaries retain personal identifiers, although these are rarely available to researchers. Identifiers may include names, addresses, dates of birth, or provincial health insurance numbers. In Canada, agencies which maintain these data (and protect the confidentiality of the data) have provided access to the data in a manner which cannot reveal the identity of individuals[66]. This is done by removing identifiers or providing only the summary tabulations required by researchers, and by maintaining some control over publication of results.

Applications of hospital morbidity databases

"Raw" databases such as the HMRI are well suited to studies of major surgery, where the actual surgical procedure should only be specified once, whether the surgical "episode" entails a single or multiple admissions. Anderson and Lomas have used this database to study trends in cesarean section in Ontario[67,68]. They were able to determine reasons for cesarean delivery from secondary diagnostic codes. They have also reported age-specific rates of coronary artery bypass surgery (CABS) in Ontario from this database in a study concerning costs of the procedure[69]. In an ecological study they estimated age-adjusted rates of CABS for 38 counties in Southern Ontario[70] and found significant between-group differences when each county was assigned (logically, but arbitrarily) to one of 5 regional CABS centres.

Halliday et al present a simple algorithm for identifying repeat admissions for the same condition from HMRI records in the province of Ontario, where health insurance (OHIP) identifiers are not unique[65]. A single, encoded, family number is present in HMRI records available to researchers. They make the assumption that only one family member of a given age and sex will be admitted for a given condition. Then the encrypted identifier, age, sex and diagnosis form the basis for searching for duplicate or multiple admissions for an individual.

The use of administrative databases for research

Only recently have administrative databases been used for analytic research. The most common, and earliest, use of a national database for etiological research in Canada has been the use of record linkage to the Canadian Mortality Database to provide mortality follow-up on large cohorts[71,72].

The utility of such databases is related to the efficiency with which they may be searched, compared to manual follow-up through traditional means (phone books, interviews, etc). There is an extensive literature on the theory and practical aspects of record linkage [73,74] and further consideration of this aspect is beyond the scope of this review.

While record linkage to vital status (alive or cause of death) has been of use for mortality follow up of occupational cohorts, administrative databases may contain rich detail on health status over a period of time and not just a single outcome (death).

Medicare data were used in the United States in the 1970s for review of various aspects of the delivery of medical care, or use of procedures[75]. The validity of Manitoba Health Services Commission (MHSC) data and its utility in the assessment of surgical procedures were reported by Roos et al in the early 1980s[75,76]. They found that the MHSC database could be used, "with appropriate caution" to investigate "serious operations and life-threatening conditions", for which diagnostic and procedure coding was likely to be accurate. An essential element of the database was the ability to create a history for an individual[77]. The history comprised not only a chronology of utilization, but data on the time during which the individual was known to the registry and thus eligible to utilize medical or hospital services.

Physician billing records provide good data on utilization[78] of health services, but may be less useful for ascertaining diagnoses. Furthermore, billing diagnoses may differ from those in clinic charts. Studney and Hakstian found that billing card diagnosis was similar to chart diagnosis for only 60% of visits to a primary care clinic in British Columbia[79]. Blinded judges found the billing diagnosis of equal or greater value in 83% of cases, although the authors concluded that any study using billing diagnosis should assess the concordance with the actual content of the clinic visit. Billing diagnoses in their study may

have been written down to "look better", as was found by their judges, but they may not reflect the true state of the patients.

Administrative databases are particularly well suited to cohort follow-up studies in jurisdictions having a good registry of residents with potential to enter the database[77]. The registry permits the determination of length of follow-up required to yield population-time for incidence rates. Hospital morbidity databases in Manitoba and Saskatchewan have been used to facilitate follow-up on cohorts. For example, Manitoba data were used to follow up a cohort of men who had a prostatectomy in the period 1974-77 for hospitalization due to recurrence of symptoms of prostatic hypertrophy[80].

Applications of drug databases

There has been considerable work done with drug utilization data in relation to post-marketing surveillance of drugs including a number of symposia on this topic. Shapiro discussed the use of epidemiological studies to conduct post-market evaluation of drugs in 1982[81]. He suggested that, at that time, the role of computerized databases was to identify cases and controls and to verify drug exposures. He warned of three limitations of record linkage of drug and medical care databases: they were primarily administrative; they seldom had information on confounders other than age and sex; and they covered a limited population size, so that the number of individuals exposed was small. He predicted that with the advent of larger computerized systems, it would be "possible to evaluate hypotheses [by means of cohort or case-control studies], between specific drugs and specific diseases in situations where confounding does not appear to be an important problem." In an issue of one drug journal devoted to adverse drug reactions, Strand discussed the potential for drug studies using the Saskatchewan databases[82]

The feasibility of record linkage of drug prescription and health care billing records has been established in several jurisdictions. Skegg linked existing prescription data to the hospital morbidity database of the Oxford record linkage study[83], although the morbidity data were collected specifically for epidemiological studies. Strom and Morse reviewed the databases which could be used to study drug utilization and morbidity with or without resorting to additional manual data extraction[84]. Included in this review were the databases of the Saskatchewan drug, medical, and hospital plans, which, although not designed for research, were deemed "a potentially very useful resource". There is a significant cost savings in conducting postmarketing surveillance in this way, given the high quality of hospital morbidity data in particular. Another major advantage over studies which include patient contact is the elimination of recall bias with respect to drugs prescribed[85]. Hospital chart review cannot include records of drugs dispensed outside of hospital.

In one Saskatchewan study all residents registered with the health plans were classified according to their exposure to non-steroidal anti-inflammatory drugs (NSAIDS) using the drug plan database[86]. Another administrative database (SHSP) was used to follow these individuals for hospitalization over a fixed time period. This permitted the estimation of incidence density (risk) of fatal upper GI haemorrhage among users and non-users of NSAIDS with no other known risk factors. While the administrative records were essential in defining population-time and for identifying probable outcome, they were not adequate to define outcome. Fatal outcomes and risk factors other than NSAIDS were ascertained by reviewing hospital charts and autopsy reports on individuals for whom a coded diagnosis was indicative of GI bleed.

Roos and Nicol demonstrated that individual histories could be compiled from the Manitoba health insurance files, even for persons for whom no claim is made[77]. A series of "snapshots" of the registration file provides information

sufficient to track individuals entering and leaving the Manitoba system and to link changes in the personal identifier (family number, date of birth and initials). In spite of changes in registration number for women as marital status changes, they were able to build complete 48-month histories on 84% of 2154 women having a hysterectomy in 1974 (2 years before and after operation). Internal and external validity of identification were excellent.

Many case-control studies have used administrative databases for the identification of cases and controls, but very few have been done using only the data available in these databases. A number of investigators have used records of large Health Service Organizations to identify cases and controls, but required chart abstraction to determine exposures[87,88]. Studies of fetal loss have used a similar strategy[89]. Hogan et al used Saskatchewan drug and physician billing records to identify women receiving isotretinoin and mailed physician (and patient!) questionnaires to the prescribing physicians[90].

Epidemiological studies can be done without access to patients or medical charts[66]. In a study of congenital anomalies, West et al used Saskatchewan drug plan records to identify a cohort of women exposed to valproate and hospital records to determine pregnancy outcomes[91]. They validated hospital records of birth with vital statistics records for the province. As an extension to this study, drug plan records of anti-convulsant prescriptions were used to identify a cohort of Saskatchewan epileptics which was linked to a pregnancy outcome registry derived solely from hospital discharge records[92].

Goldacre et al reported on two studies of cardiovascular disease and vasectomy based on hospital records[93]. One of their studies used case-control methodology. Cases were identified from hospital and mortality records of myocardial infarction using the Oxford Record Linkage Study data. Hospital and death controls were selected in a similar fashion and with matching criteria. Exposure (vasectomy) was determined by linking to files maintained by the

vasectomy clinic, which provided exclusive surgery in the region. They also used cohort methodology and record linkage of the Scottish hospital discharge database to follow up a large number of men having vasectomy, meniscectomy or nasal surgery.

Ray et al conducted a case-control study of thiazide use and hip fracture using only the Saskatchewan hospital and drug plan databases[94]. Their case definition for hip fracture excluded those with a concurrent diagnosis of malignant neoplasm or major trauma; chart review of a sub-sample validated the case definition which had been based on ICD coding. Cohen and Hammarstrand have studied the relationship between timing of Papanicolaou testing and cervical cancer using the Manitoba physicians' claims database. They report on longitudinal data of a population sample[95] and have also used this data in a case-control study design[96].

Limitations of administrative databases for epidemiological research

Biases may limit the validity of case-control studies which use only administrative data. Perhaps most important is the absence of data on confounders because it is not possible to predict the direction of such bias, if it occurs. Information bias may arise because of migration, or failure of the database to include data relating to exposures in the remote past. Exposures may be incompletely represented in the database. There may be selection bias if the referent population is not uniformly included in the database.

Successful studies have been done when the exposures and outcomes are conditions or treatments which are completely represented in the data sets. Potential confounders or matching variables were limited to those which are found in the data; for example, age, hospital, admission date and residence in studies of cardiovascular disease and vasectomy by Goldacre et al[93].

As mentioned above, the provinces of Manitoba and Saskatchewan are ideal settings for research using administrative data. Other provinces have the capability of linking records via health care numbers, but do not have a longitudinal registry which will permit determination of person-time denominators. Quebec is developing a policy for dissemination of their MED-ECHO data, which comprises all hospital separations in the province. L'Abbé et al used physician billing data in Quebec to determine changes in physician visits in relation to the installation of urea-formaldehyde foam insulation in homes, but they were unable to assess diagnostic codes[97]. Other provinces use HMRI to process hospital separation records, but these are complete (in terms of hospitals represented in the data) only for Ontario prior to 1988. Saskatchewan is the only province with a universal drug plan which permits population-based pharmacoepidemiology studies.

CHAPTER II

Methods

The Saskatchewan Hospital Services Plan data were used to estimate relative risk of ectopic implantation, conditional on fertilization, for a number of potential risk factors. Identification of cases and controls was on the basis of primary separation diagnosis. An **index separation** is that hospital separation which was the basis for selection of a case or control (Appendix C).

Case-defining separation

All cases of ectopic pregnancy treated and discharged from hospital in Saskatchewan during April, 1984 through March, 1987 were included. Each ectopic pregnancy episode was considered a case. For analyses based on persons, rather than pregnancies, only the most recent ectopic pregnancy episode for an individual was included.

Control-defining separation

A control was chosen for each incident EP by selecting a separation record for a woman in the same 5-year age group who had any pregnancy-related hospitalization on or following the date of the EP discharge. Sampling of controls from the SHSP database was on the basis of primary diagnostic code. Eligible codes were those which indicated any pregnancy condition other than ectopic pregnancy. A case individual was eligible to be selected as a control on the basis of a normal pregnancy occurring in the same time period. In this instance the individual was included as a case or control as of the date of the respective index separation. Control selection was not at random, but

was dependent on the chronological order in which SHSP records are stored on magnetic tape. Analyses based on persons included only the control matching the most recent case (index) separation.

Raw data description and cleanup

Exposure data were extracted from all Saskatchewan hospital separations, for cases and controls, dating back to 1970. Anonymized data were provided by the SHSP, on magnetic tape, in a specified format and subject to a number of conditions on the publication of results (Appendix A). Two sets of data were received from SHSP. The first contained records for cases and controls for the fiscal years 1984/85 and 1985/86; the second, for 1986/87. The raw data consisted of a series of separations records for each individual. Each record in a series corresponds to a single hospital stay for that individual, and contains the information shown in Appendix B. Two diagnostic codes and the single, major, procedure code were available. The two data sets were extracted by a record linkage program devised by SHSP. This program generated sequential identification numbers to replace the SHSP internal unique identifiers. The data extraction for the third year was done independently of the first two years. Numbering began with one in each data set, so that a person identified in both linkage runs would have different identifiers in the two data sets.

The raw data were transferred from a mainframe computing facility to the LCDC communications computer, thence to microcomputer. Data transfers were done using high-speed lines and with error-detecting protocols. No invalid ASCII characters were found in the data received by the microcomputer. All data cleanup and transformation was done on the microcomputer. When necessary, transformed files were transferred to the University of Ottawa

mainframe computer using an error-detecting protocol (KERMIT*). All data fields were first checked by scanning frequency tabulations.

Diagnoses were coded to the 8th and 9th revisions of the International Classification of Diseases (ICD)[98,99]. Procedure codes were coded to the 8th revision of the ICD or to the Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures (CCP)[100]. Alphanumeric diagnostic and procedures codes were converted to unique 4-digit codes with an implied decimal point to simplify data transformations. All ICD "V" codes were converted to negative numbers; for example, V25.2 was originally coded as V252 and became -252. The "&" and "A" prefixes were removed from ICDA-8 diagnostic codes and CCP procedure codes, respectively. Trailing decimal zeros were absent from diagnostic and procedure codes. For example, the ICD-9 code for Delivery, uncomplicated is 650 and appears in the data field as "650 ". Many procedure codes were specified to two or three digits and trailing zeroes were added to ensure uniqueness. Because procedures were coded with leading zeroes and were left-justified in the data field, this was a correct translation of the codes.

Search for individuals appearing in both data sets

Because two independent data extractions had been done, a woman having multiple pregnancies could be included in both data sets. For analyses based on individuals, it was necessary to search for duplicate records to ensure that an individual's history would not be counted more than once (Figure 13). Based on the separation date and date of birth, a search was made for ectopic pregnancy separations which may have been duplicated in the two data sets. This would occur if a case from 1984 or 1985 had a case-defining or control-defining separation in 1986. The historical record of ectopic pregnancy from the

* KERMIT is a commonly used error-checking data transfer protocol devised at Columbia University and implemented for many computer operating systems.

1986 data would match that from the earlier data set. Duplicates were verified by examining historical records to ensure a complete match. To facilitate analysis by episode, a record for an incident EP other than the most recent was given a new unique ID and was grouped with a copy of all prior separation records for that individual.

Transformation of data to one record per episode

I wrote a program to transform raw data for an EP episode and all prior separations into a single summary record. The records were first ordered by unique ID and within that, by date of separation, the most recent coming first. Certain fields were selected from the record(s) of the index separation and appended to these were a number of historical data derived from earlier records of hospitalization (Appendix B).

The transformation program was written in Microsoft Fortran to run on an MS/DOS compatible microcomputer. It can search records for the occurrence of up to 20 conditions of interest. Input of ICD-9, ICDA-8, and CCP codes follows a generalized format which allows recoding of multiple individual codes or ranges of codes to a single target condition. Each target was coded to signify whether it implied pregnancy (gravidity of 1) or pregnancy outcome (parity of 1) (Appendix C).

The program has built-in rules for identifying clusters of separations which are probably transfers or multiple hospital admissions for one episode of a condition, and treats these as a single hospital episode. Critical to this is the determination of elapsed days (delta) from prior separation to new admission. A probable transfer is detected when one of the following is true:

- delta ≤ 1 day and prior separation type = 5 (transfer to another hospital)
- delta ≤ 14 days and the later separation is for ectopic pregnancy
- delta ≤ 7 days and both separations recode to the same, nonzero, condition of interest

All such instances were manually verified. A warning was printed when multiple conditions of interest were detected or when the total gravidity from the cluster of separations was greater than 1. The latter suggests that there is a conflict between diagnoses, such as the instance where a provisional diagnosis of spontaneous abortion was made at one institution and a definitive diagnosis of ectopic pregnancy was made following transfer to another. The "transfer" rules were arrived at during the process of revising and debugging the transformation program.

There is potential for adapting the program to transform other similar data sets and conditions, for expanding the number of conditions of interest, or for using the gravidity/parity counting for other purposes. Transformations of four sets of source codes are handled at present: ICD-9 diagnosis, ICDA-8 diagnosis, CCP procedures, and ICDA-8 procedures (Appendix C).

The fields copied from the index separation record were unique ID, case status, SHSP age group, primary & secondary ICD-9 diagnostic codes, operation code, date of separation, admission & separation types, days to surgery, and days in hospital. The matching case ID and the sequence of case-defining separations (for cases having multiple EP) were added to the record later. This ensured that matched analysis of episodes or individuals was possible. Data derived from the index record were age in whole years and recoded diagnosis and operation codes. ICD-9 and CCP codes of interest were recoded to 14 categories representing delivery and plausible surgical or diagnostic risk factors. The historical derived data included all separations for an individual except the index separation: age at first known separation;

number of separations excluding transfers; gravidity including index pregnancy; parity excluding index pregnancy; age at first and last known prior pregnancy; number of separations and age at first and last known prior separation for 13 conditions of interest (Table 2). Presence of an IUD at the time of the index pregnancy was determined from the secondary diagnostic and procedure codes on the index separation.

Sufficient output was printed by the transformation program to identify some errors in the raw data: year or month of birth for an individual different from the index separation on a prior record, diagnostic incompatibilities within the "transfer" cluster of separations, wrong order to data. Output from the program's diagnostic log file were used in the process of revision of the program itself and the transformation of ICD codes to conditions of interest.

Table 2. Conditions of interest from historical records (1970-87)

Acute salpingitis
Chronic salpingitis
STD or other genital infection
Endometriosis
Induced abortion
Infertility
Sterilization
Tubal repair
Ectopic pregnancy
Abortion, spontaneous
Tubal surgery
Caesarean section
Delivery

Internal and external validity of SHSP data

Validation of individual histories

A number of approaches to validation of the histories were taken. There were three concerns. First, were the data for an individual internally consistent? Second, was there selection bias in the ascertainment of a case history compared to that of a control history? Lastly, was it possible to test the external validity of histories?

The internal consistency of the coding for ectopic pregnancy was assessed by examination of the secondary diagnostic and procedure codes on the index records. No attempt was made to review hospital records. In the (majority) cases of emergency treatment for EP, the secondary diagnosis could relate to a pre-existing medical problem, complications of surgery for removal of the EP, or additional findings at surgery. The most likely misclassifications would be errors in coding other pregnancy outcomes as EP, as happened in a Prince Edward Island hospital in the late 1970s[101]. Beral examined procedure codes associated with EP separations and found some systematic miscoding of induced abortions to abdominal pregnancy (ICDA-8 codes 631.0 and 631.4) in England[89].

Possible invalid, or invalidating, secondary diagnoses are conditions relating to pregnancy beyond the first trimester or to an intrauterine pregnancy; for example, vaginal or caesarean delivery. Because a single procedure code was available in the SHSP data, it was anticipated that in the majority of cases this would refer to the definitive treatment procedure, that is, surgical removal of the EP. Related diagnostic procedures, or those which code complications such as acute blood loss support the diagnosis of EP. Invalid, or invalidating,

procedure codes would include any indicating term delivery or others which could not conceivably be undertaken at the time of treatment for EP.

Delivery or abortion outcomes were validated by examination of procedure codes on the index separations for invalidating procedures which could not possibly be related to pregnancy. For primary diagnoses which did not indicate a pregnancy outcome, secondary diagnostic and procedure codes were examined in order to classify, where possible, the outcome as abortive or delivery. No forward-linked (that is, following the index separation) records were available which could have provided outcomes for separations that indicated continuing intrauterine pregnancy.

Coding of prior pregnancy-related separations were examined in the same fashion. Tabulations of diagnostic and procedure codes from historical records were reviewed to identify any prior hysterectomy, which would be inconsistent with case or control selection.

Having unique identifiers does not ensure an exact match in a record linkage study. The SHSP database has already had consistency checks applied, and the processing of records by HMRI includes verification of coded data by the hospitals. The one variable which could provide an internal check on the linkage process was date of birth (DOB). DOB on all historical records was compared to DOB found on the index separation. Year and month, but not day, of birth were known. Records were accepted if either year or month agreed.

Internal validity

The transformation procedure derived four variables which summarized the known history of each individual, in addition to the age at first known and last known prior conditions of interest. The four summary variables were years between first known separation and index separation, number of hospital

episodes on record, gravidity and parity. Gravidity and parity are of interest as potential confounders.

Internal validity - distribution of days between separations.

The interval between hospitalization episodes for an individual can be thought of as a "time-to-failure". In theory, time-to-failure for a Poisson variable, such as the number of hospitalizations, arises from a stochastic process having an exponential distribution[102]. Thus, it should be possible to model the distribution of intervals in a meaningful way. Estimation of model parameters utilizes all of the time-related data available and is potentially a powerful means of assessing the validity of the process used to accumulate individual histories. If the data fit a reasonable distribution, then there is some assurance that the observed histories of hospitalization represent reality. It also provides a basis for parameterization of the distribution of intervals for group comparisons.

The records for an individual were scanned by another Fortran program to generate a series of separation-to-separation intervals in whole days greater than zero. The time between consecutive hospitalization episodes for each individual was calculated as the number of days between separation dates (delta). Transfer separations were ignored using the same algorithm as for data transformation. Also ignored were any zero-time intervals which were presumed to represent a single hospitalization episode. Case or control status was based on status at the index separation. The two cases which were also selected as controls were included in both groups. For a patient having n separations, there would be up to n-1 intervals defined. The derived data set has one record per interval which contained ID, date of the earlier of the (2 adjacent) separations, case status, separation number (index separation being 0) and time in days to the next separation.

Descriptive statistics, including mean and standard deviation, were computed for the complete set of intervals. The frequency and cumulative frequency distributions of intervals were compared to normal, log normal, exponential and general gamma probability distributions. Parameter estimates for the best fit to a distribution were derived using maximum likelihood estimation or non-linear regression. The estimates were obtained using a cumulative frequency distribution for days between separations as input to non-linear regression by BMDP3R[103]. The interval for the frequency distribution was one day. Only non-zero intervals were input. Days were transformed to years and frequency was transformed to cumulative proportion because BMDP3R could not converge with the larger values. Parameters were estimated for subsets of data as well, including cases only, controls only, patients with specific histories such as PID, and patients dichotomized on some other variable such as HISTORY or number of separations.

External validity

A test of validity using an external source of data was to compare observed parity with estimated parity at birth as reported by physicians' notices of birth (PNOB). Parity calculated by the translation program did not include the index separation. True parity includes the index delivery and corresponds to birth order as reported by PNOB in Canada.

The estimated parity is a lower limit, since there may have been, for any individual, a delivery that never required hospitalization, or a delivery occurring outside the province of Saskatchewan. Given this limitation, the known parity of controls having a live birth delivery outcome was compared to vital statistics records of parity for the general population.

The data used were calculated parity for all three years for controls having an index separation coded as delivery (conditions 13 or 14). The index

delivery would therefore be the last one occurring in the data set for an individual. Published data on birth order by maternal age group for Canada, 1986, were chosen to represent the distribution of parity in the population[104]. Age specific data were required because the age distribution of controls was different from that of all delivering women. Data for Saskatchewan would have been preferred, but these were not readily available. The vital statistics data were used to predict the total number of deliveries among controls at each birth order as follows:

Let n_{pk} be the predicted number of controls at parity p in age group k. Then

$$n_{pk} = N_{pk} \times a_k / A_k$$

where N_{pk} is number at parity p, age group k, Canada

A_k is number in age group k, Canada

a_k is number in age group k, Saskatchewan

Let O_p be the observed number of controls at parity p. Assume that the fertility among Saskatchewan women is the same as among all Canadian women. Then the predicted total number of women at parity p is

$$E_p = \sum_k n_{pk}$$

The distribution of E_p and O_p were compared using a Chi-square test. Because of small predicted numbers at high parity, totals for parity 4 or greater were pooled.

Power

The expected sample size of approximately 630 cases was used to estimate power prior to data acquisition. The inverse of the Casagrande & Pike approximation for sample size estimation was used to estimate power for various combinations of case and population prevalence of salpingitis[105]. Based

on hospital morbidity data for Saskatchewan, and under the assumption that half of women having PID will not conceive, either by choice or because of infertility, it was estimated that the background prevalence of hospitalization for any prior episode of PID in conceiving women could be as high as 0.88%. Prevalence rates for salpingitis due to *Chlamydia trachomatis* among cases of ectopic pregnancy may be as high as 50%[106,107,108]. It has been suggested that 10% to 25% of episodes of PID require hospitalization[109,110]. Typical odds ratios for salpingitis in case-control studies of ectopic pregnancy are of the order 4-6[7]. Power curves for varying assumptions of case and population prevalence are shown in Figure 4. Corresponding odds ratios are shown in Figure 5. Assume that the prevalence of ever having salpingitis requiring hospitalization among pregnant women is 1%. With 630 cases and controls, the study has power 98% to detect an odds ratio of 5.2 or greater if the prevalence of hospitalized salpingitis among cases were 5% (α=.05, 25% of PID episodes hospitalized, PID prevalence among cases is 20%). Power to detect an odds ratio of 4.7, corresponding to a 4% prevalence among cases is 90% (α=.05, 10% of PID hospitalized, PID prevalence among cases is 40%).

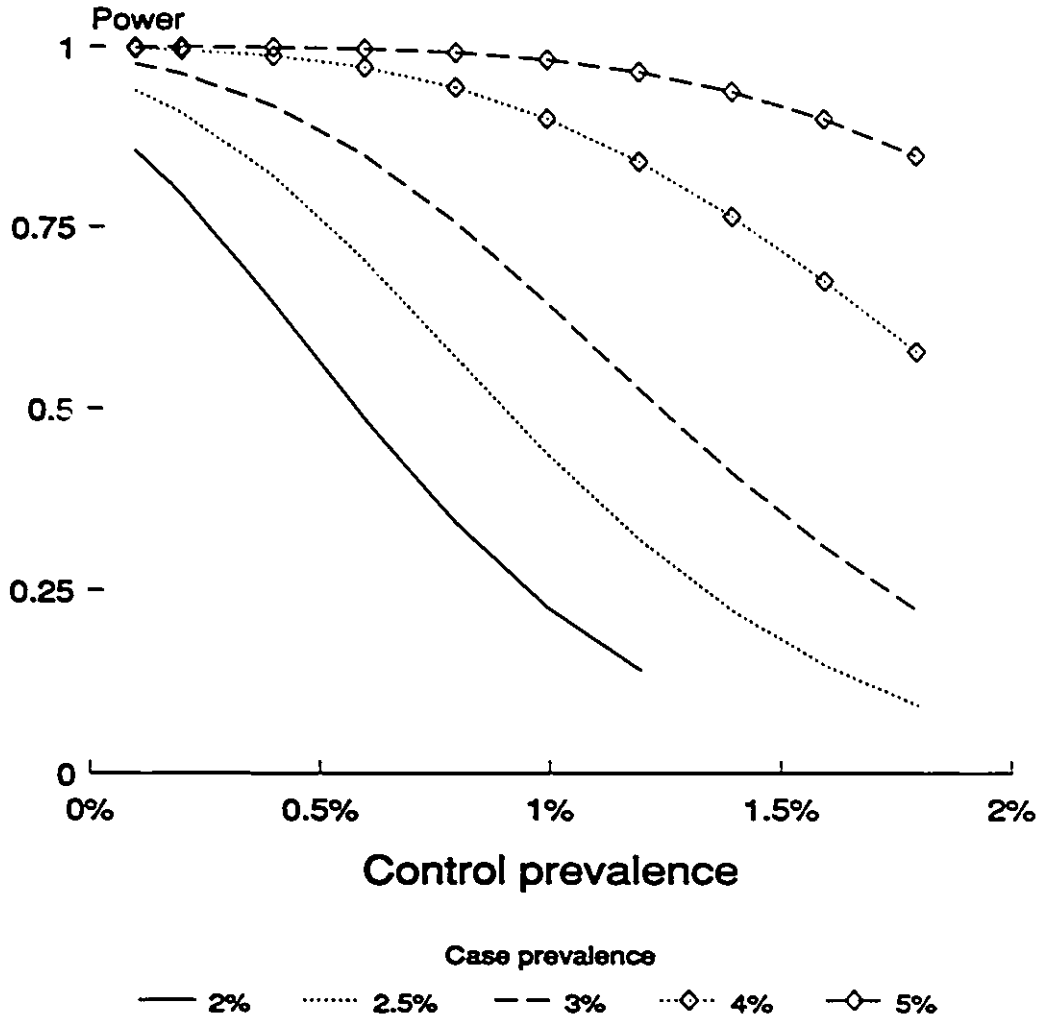

Figure 4. Power of case-control study with 630 cases and controls

Curves are based on the inverse of the Casagrande-Pike approximations for sample size determination (α=.05, 1 tail).

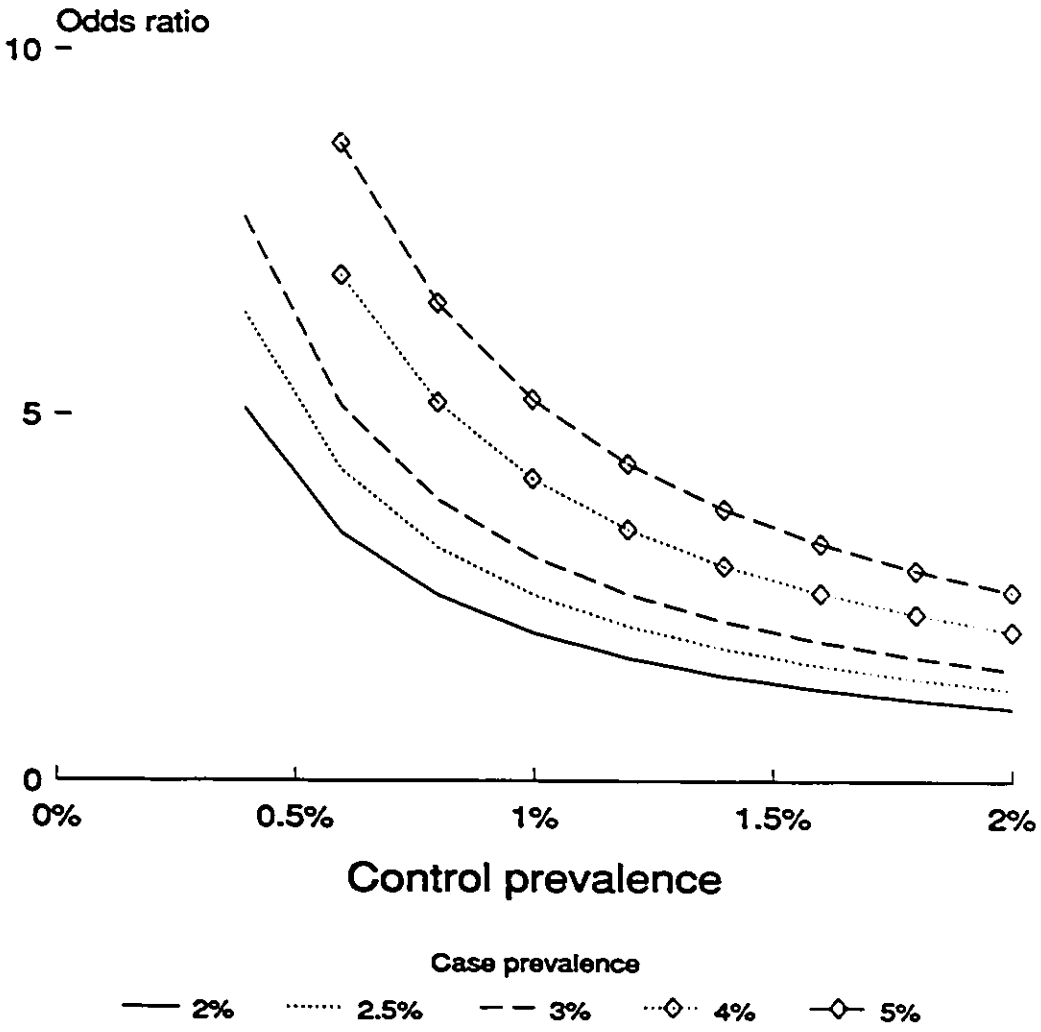

Figure 5. Odds ratios estimable from case-control study

Curves shown correspond to the power curves in Figure 4.

Case-control data analysis

The study was a retrospective, case-control design with 1:1 matching on 5-year age group and time of separation. The unit of analysis was either a pregnancy (all case and control index separations included) or an individual (only the most recent index separation included). Time of separation was matched only as a convenience and to ensure that control separations at least fell within the same fiscal year as the case separations. The control sample was matched on age to provide sufficient controls in the older age groups, at minimal (1:1 ratio) cost.

The data permitted description of ectopic pregnancies with respect to age, admission status, outcome, and surgical procedures. The transformed data were used, either by single year of index separation, or for all three years combined. Creation of the combined file is described on page 36. The files were imported into a format used by Epi Info[111] or SPSS-PC+[112]. Condition variables were collapsed into dichotomous variables representing the presence or absence of the condition in a patient's history. Additional variables were derived to represent the presence or absence of one or more potential risk factors.

Mean values for variables used to describe episodes were compared by the paired t-test or a two-sample Kolmogorov-Smirnov test, using SPSS/PC+. Some comparisons were made for each of the three years.

The data were suitable for estimation of odds ratios for EP based on individuals or episodes. For the estimation of etiologic fraction, the more appropriate unit of analysis was an episode of EP, because one wishes to attribute ectopic pregnancies to various factors. This is also consistent with the proposed nature of risk estimates, that is, that they be conditional on conception. All other case-control studies of EP have used the individual as the

unit of analysis; either approach may be justified when one wishes to estimate the risk of ectopic implantation for a subsequent pregnancy in an individual. Many investigators have excluded individuals having a prior ectopic pregnancy in order to avoid confounding of other factors with prior EP. This also eliminates the opportunity to use multiple episodes of EP.

In order to assess interactions among risk factors, one is forced to use either a stratified analysis (ignoring pair matching) or conditional logistic regression. Schlesselman suggests using stratified analysis, the method chosen for this study[113]. Stratification by age group, rather than pair matching, can be considered appropriate in this study because the one-to-one selection of controls was a convenience for data extraction and done for the sake of efficiency.

It can be argued that the control group was a stratified random sample of pregnant women having a hospital stay, where the selection was based on the random occurrence of a hospitalization for another woman's ectopic pregnancy. Since the size of the control population in each age stratum was not known directly, the controls may be considered to have been frequency matched by age and time of index separation[114]. If one considers that controls selected in the manner described share some other, unmeasured, characteristics, then stratified analysis would be inappropriate. However, such is not the case. Therefore, a stratified analysis is appropriate. Both matched and stratified analyses were done.

Odds ratios were computed for matched individuals or episodes. Test-based 95% confidence intervals were estimated by Fleiss' method[115]. Mantel-Haenszel odds ratios, stratified for age group, year of index separation, prior EP, or other variables were estimated using the Epi Info programs Analysis and Statcalc[111]. These programs also provide test-based 95% confidence intervals by the method of Robins, Greenland, and Breslow[116], and exact confidence

intervals if requested[117]. Stratification by year of index separation was done in order to assess stability of estimates over time.

Multiple logistic regression was used to estimate odds ratios, conditional on multiple factors[118]. Age stratum was included in the logistic model through a series of 5 dummy variables. The lowest age strata (<15, 15-19) were combined, as were the highest (40-44, ≥45). Unconditional logistic regression was used because the stratum sample sizes for most strata were large. In such circumstances there should be little bias in the unconditional regression parameter estimates[119]. Bonferroni confidence intervals for regression parameters were used to estimate approximate joint confidence intervals for odds ratios in the final logistic regression model[120]. Approximate 95% confidence intervals were based on the one-sided normal deviate having probability 0.025/g, where g is the number of estimated parameters in the model.

Because the incidence of EP is low in the population studied (women conceiving), etiologic fractions should be estimable[121]. Recall that risk factors are represented in the data by hospitalizations for the conditions. For those conditions which do not always cause a hospital stay, control prevalence will underestimate population prevalence. Surgical conditions should be well represented, while as few as 10% of salpingitis episodes require hospitalization[109]. Etiologic fractions (and approximate 95% confidence intervals) for single risk factors were estimated by the method described by Schlesselman for matched* or stratified data[122]. Etiologic fractions (attributable risk) estimated in this way were not adjusted for other factors, although an independent adjustment for prior EP was done. Risk factors measurable in this study are

---

* "With matched studies, one may disregard the pairing and stratify cases and controls into subgroups on the basis of the matching variables." (Schlesselman, p224)

correlated, and it is clear that EP has a multifactorial etiology. Adjustment for multiple factors is feasible, but even the crude etiologic fractions may be used to compare the relative importance of risk factors[123]. Because it is known *a priori* that risk factors are not independent, the etiologic fractions are not additive. For example, prior ectopic pregnancy confounds all other factors.

CHAPTER III

Results

Internal consistency of hospital data

Coding of primary diagnosis of ectopic pregnancy

The combined data sets included 5841 records representing hospital separations for 1324 individuals. Seventeen cases appeared in both the 1984-85 data set and the 1986 data set. Two of these appeared as controls in 1986/87, having had a pregnancy with normal implantation in that year.

Only one case had a secondary diagnostic code (Tables 3,4,5) which was incompatible with ectopic pregnancy: 663.1 (umbilical cord around neck, with compression). History records for this case indicated Gravida 4, Para 1, Aborta 2 with Shirodkar suture having been placed 2 months prior to index hospitalization. The index operative code indicates D&C. This episode likely represents a spontaneous abortion since Shirodkar sutures are normally placed after the first trimester, and only in the presence of a confirmed intrauterine pregnancy.

Eighty per cent of the 677 ectopic pregnancy separations for the three years were coded with a definitive surgical procedure (Table 6). The majority (62%) were coded as either 78.52 (partial salpingectomy with removal of tubal pregnancy) or 86.30 (removal of intraperitoneal embryo following rupture of tube), the only CCP codes specific to tubal pregnancy. Forty-five (6.7%) had no procedure listed. Two cases had procedure code 85.70 (episiotomy); both had elective admissions and likely represent misclassification of the primary

diagnosis. No other inconsistent procedure codes were found. When other compatible codes were included, 573 episodes had surgical procedures which could have been done at the time of treatment for EP (codes 77.12 through 80.30; 86.30; or 86.49).

The three individuals with inconsistent coding were excluded from further analysis.

Table 3. Secondary diagnostic codes on case separations by ICD-9 chapter

Number	ICD-9 Chapter and selected diagnoses (n)
3	I. Infectious disease
8	II. Neoplasm
	uterine leiomyoma (6)
	benign ovarian (2)
5	III. Endocrine, nutritional, metabolic
	obesity (4)
15	IV. Diseases of blood
	acute posthaemorrhagic anemia (11)
	anemia, unspecified (4)
2	V. Mental disorders
5	VI. Diseases of the central nervous system
5	VII. Diseases of the circulatory system
	haemorrhage, unspecified (3)
5	VIII. Diseases of the respiratory system
23	IX. Diseases of the digestive system
	appendicitis without peritonitis (1)
	peritonitis (5)
	peritoneal adhesions (14)
147	X. Diseases of the genitourinary system
	(see Table 4)
17	XI. Diseases of pregnancy
	(see Table 5)
1	XII. Diseases of the skin and subcutaneous tissue
1	XIII. Diseases of musculoskeletal system
9	XIV. Congenital anomalies
	anomaly of fallopian tube or broad ligament (8)
12	XVI. Symptoms, signs and ill-defined conditions
	shock (3)
13	XVII. Injury and poisoning
	surgical injury (6)
	surgical complication (7)
41	Supplementary (V-) codes
	sterilization (16)
	insertion or removal of IUD (1)
	other contraceptive management (5)
	tuboplasty (1)
	other procreative management (4)
	presence of IUD (2)
354	No secondary diagnostic code

Table 4. Secondary diagnostic codes on case separations (ICD-9 Chapter X)

Number	ICD-9	Diagnosis
17	599.0	urinary tract infection
1	611.7	signs/symptoms in breast
1	614.0	acute salpingitis
8	614.1	chronic salpingitis
2	614.2	salpingitis, acute or chronic
7	614.4	chronic or unspecified parametritis
1	614.5	pelvic peritonitis
32	614.6	pelvic peritoneal adhesions
3	614.9	unspecified pelvic inflammatory disease
2	615.1	chronic uterine inflammatory disease
6	616.0	cervicitis, endocervicitis
1	616.3	abscess of Bartholin's gland
8	617	endometriosis
1	618.0	prolapse of vaginal walls
3	620.0	follicular cyst of ovary
25	620.1	corpus luteum or haematoma of ovary
17	620.2	unspecified ovarian cyst
6	620.8	other disease of ovary, tube, broad ligament
3	621.6	malposition of uterus
1	623.8	other disease of vagina
6	628.2	infertility of tubal origin
5	628.9	infertility of unspecified origin
4	629.8	other disorder of genital organs

Table 5. Secondary diagnostic codes on case index separations (ICD-9 Chapter XI)

Number	ICD-9	Diagnosis
3	633	ectopic pregnancy
1	634	spontaneous abortion
5	639	complications of ectopic pregnancy (haemorrhage, laceration, shock)
2	646	other complications of pregnancy
2	648.2	anemia complicating pregnancy
3	656.1	Rh immunization
1	663.1	**umbilical cord around neck, with compression**
1	669.1	obstetric shock

Table 6. Procedures coded on all case ectopic pregnancy separations

Total	CCP*	Procedure
1	02.28	Chest x-ray
4	02.88	Ultrasound of gravid uterus
1	02.89	Ultrasound site not specified
1	03.26	Gynaecological examination
1	11.71	Removal of IUD
7	66.19	Laparotomy
26	66.83	Laparoscopy
3	77.12	Wedge resection of ovary
4	77.19	Other local excision of ovary
1	77.20†	Unilateral oophorectomy
33	77.30†	Unilateral salpingectomy-oophorectomy
1	77.70	Freeing of adhesions of ovary, tube
1	77.82	Other biopsy of ovary
21	78.00†	Salpingotomy
41	78.10†	Total salpingectomy
4	78.21†	Total bilateral salpingectomy - removal of both tubes
1	78.22†	- removal of remaining tube
1	78.42†	Bilateral ligation and division of fallopian tubes
1	78.49†	Other bilateral destruction or occlusion of tubes
2	78.51†	Other salpingectomy - excision or destruction of lesion of tube
389	78.52†	- salpingectomy (partial) with removal of tubal pregnancy
2	78.53†	- bilateral partial salpingectomy, unqualified
7	78.59†	- other partial salpingectomy
5	78.61	Repair - suture of tube
12	78.69	-other repair
2	78.91†	Aspiration of tube
9	78.99	Other operation on tube, NEC
1	80.19†	Excision or destruction of a lesion of the uterus
3	80.30†	Total abdominal hysterectomy
3	81.01	D&C following delivery or abortion
5	81.09	Other D&C
1	81.29	Other excision or destruction of lesion or tissue of uterine supports
4	82.00	Culdocentesis
1	82.22	Excision or destruction of lesion of cul-de-sac
2	85.70	**Episiotomy**
28	86.30†	Removal of intraperitoneal embryo
6	86.49†	Other removal of embryo, NEC
45		no code

*Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures
†procedure may be used to treat ectopic pregnancy (539/677)

Coding of pregnancy diagnosis for controls

Outcome could not be classified on the basis of diagnostic and procedures codes for 205 (29%) of 687 control index pregnancies (Table 7). Of these 205, 37 had procedures which indicated labour or delivery, 28 had procedures suggestive of pregnancy, 6 had procedures compatible with pregnancy, and the remaining 134 had no procedure coded. These proportions were consistent over the three years.

Table 7. Pregnancy status for control index separations

Number	Status
413	Delivery
-	Stillbirth (no stillbirth codes recorded)
50	Spontaneous abortion
5	Induced abortion
14	Abortion (unspecified)
205	Pregnancy in progress

Other consistency checks

Historical records were compared for the individuals who appeared in both raw data sets and were in all instances exactly the same.

For the years 1984/85 and 1985/86, the data acquired in the first linkage run, date of birth did not agree on 6 (0.2%) of 2965 historical records, but only the month differed (8 instead of 3 on 3 records for one individual; 5 instead of 10 on 3 records of another). No records were rejected from that data set.

For the year 1986/87, however, many inconsistencies were found. For 60 (3.8%) of 1566 historical records, date of birth differed from the index

separation for 29 individuals. Only two records were rejected because both year and month differed. There was no pattern to the differences. Most occurred in the years prior to 1986/87, that is in records which had already passed SHSP validity checks. It was not a matter of the index separation being "in error". For one case, 18 of 39 historical records had birth date 51/7 instead of 51/8.

There were no coded hysterectomy procedures among the historical records for any case or control.

The age group codes provided in the raw data were computed by subtracting year of birth from year of separation. For some individuals this classification differed by one from the age group as computed from year and month of birth and date of separation (Table 8). Because control selection had been on the basis of the SHSP-derived age group, age was considered a *potential* confounder, even though it was a matching variable. In no case-control pair was the difference in ages 6 years or greater.

Table 8. Disagreement between age group codes and computed age group

	Age group* of control assigned by SHSP		
Age group of case*	Lower group	Same group	Higher group
15-19	2	27	6
20-24	7	122	20
25-29	25	202	23
30-34	24	135	5
35-39	8	52	1
40-44	5	8	1

* computed from year, month of birth and date of separation

Estimation of parameters for models of distribution of days between separations

The expected cumulative probability distribution for a simple exponential waiting time:

$$F(x;\beta) = 1 - e^{-x/\beta}$$

did not fit the data well. The frequency distribution of waiting times for all the data was bimodal (Figure 6), so a compartmental model was chosen to represent waiting times drawn from independent distributions. The rationale for doing this is that there may be a group of conditions which cause multiple hospitalizations over a short period and be unlike the majority of conditions affecting women of this age. One example is a healthy woman with a complicated pregnancy, which may require several admissions over a 9-month period. Recall that the times between transfers, as defined in Chapter II, page 37, were excluded.

The two compartment model has the cumulative probability distribution:

$$F(x;\beta_1,\beta_2) = 1 - pe^{-x/\beta_1} + (1-p)e^{-x/\beta_2}$$

where β_i is the mean waiting time for compartment i,

p is the proportion of waiting times which derive from the exponential distribution with mean β_1

The maximum likelihood estimator for β in the single parameter model is the arithmetic mean waiting time. Estimates of β_i for the two parameter model were determined by non-linear regression of the observed cumulative frequency with time in days as the independent variable (Figure 7). Note that the single parameter model is equivalent to forcing p=1.

The nonlinear regression results for all data, and various subsets designed to assess the stability of the model were comparable. The comparability of case and control histories was assessed, in part, by the agreement with the model for all data.

The 2 compartment model shows stability over the three years of cases (Table 9). As might be predicted, the proportion of long intervals is increased and the long-interval mean is greatest for the most recent data; longer histories are possible since the data set is truncated below 1970.

Data for cases or controls alone also fit the model well with little difference between corresponding parameters for the two groups (Table 10, Figure 8). The distributions are significantly different (by Kolmogorov-Smirnov 2-sample tests) because of the large sample sizes.

Table 9. Parameter estimates for exponential models of time between separations - by year of index separation

Parameter	All data	All data	1984	1985	1986
n	4248	4248	1431	1325	1492
β_1^{-1} (year^{-1})	0.72	0.56	0.59	0.55	0.55
β_1 (years)	1.38	1.78	1.69	1.81	1.82
$s(\beta_1^{-1})$	0.0035	0.0007	0.0014	0.0008	0.0013
p	1	0.813	0.784	0.814	0.842
β_2^{-1} (year^{-1})		16.7	12.4	18.1	22.4
β_2 (days)		21.9	29.4	20.2	16.3
$s(\beta_2^{-1})$		0.276	0.223	0.275	0.777

Table 10. Parameter estimates for exponential models of time between separations - by case status

Parameter	All data	Cases	Controls
n	4248	2058	2190
β_1^{-1} (year^{-1})	0.56	0.53	0.60
β_1 (years)	1.78	1.89	1.67
$s(\beta_1^{-1})$	0.0007	0.0007	0.0010
p	0.813	0.838	0.793
β_2^{-1} (year^{-1})	16.7	17.7	16.7
β_2 (days)	21.9	20.7	21.9
$s(\beta_2^{-1})$	0.274	0.343	0.29

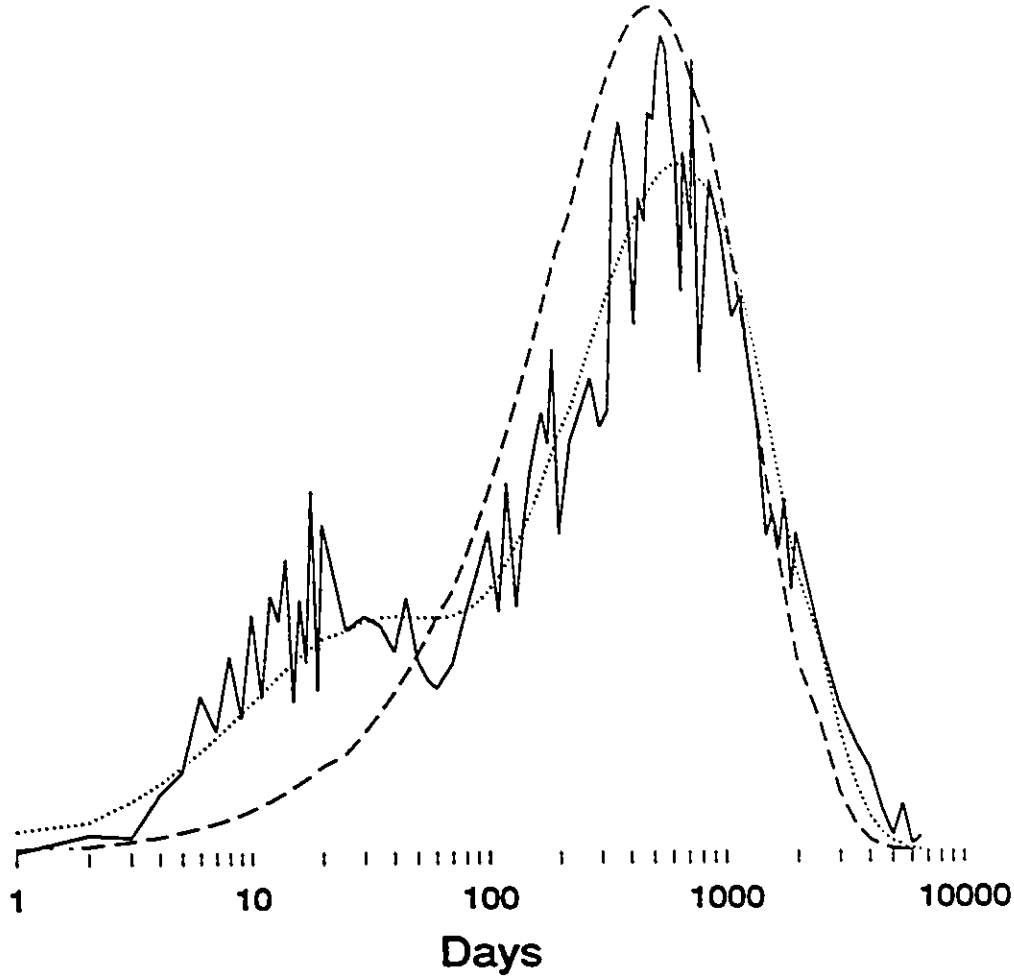

Figure 6. Frequency distribution of waiting times between separations

Solid line is observed frequency density; dotted line is probability density, based on differences in the 2 compartment cumulative distribution; dashed line is probability density for a simple waiting time distribution. Density is normalized for logarithmic scale on the time axis so that area under each curve is 1.

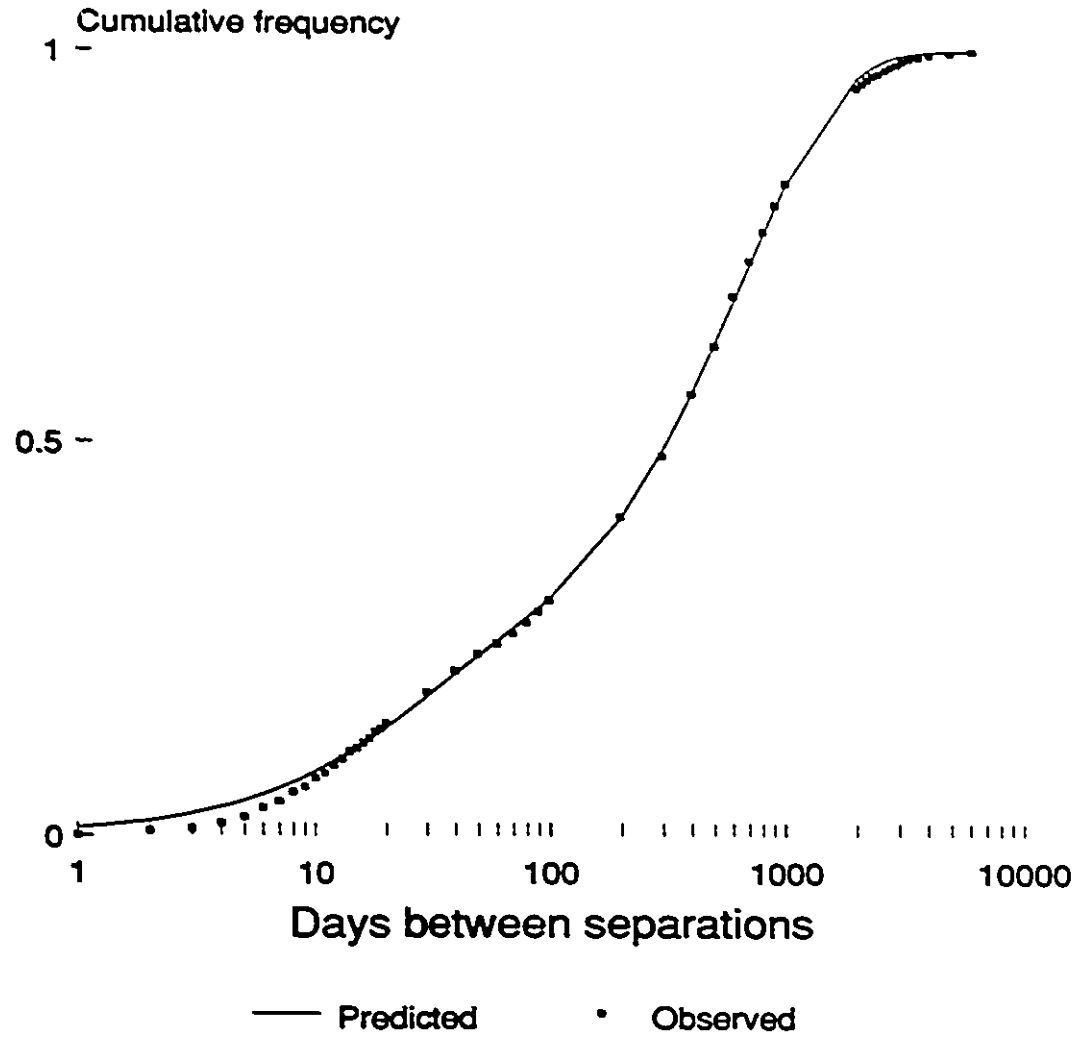

Figure 7. Cumulative distribution of waiting times between separations

Square box is observed cumulative distribution; solid line is predicted distribution based on the 2 compartment model.

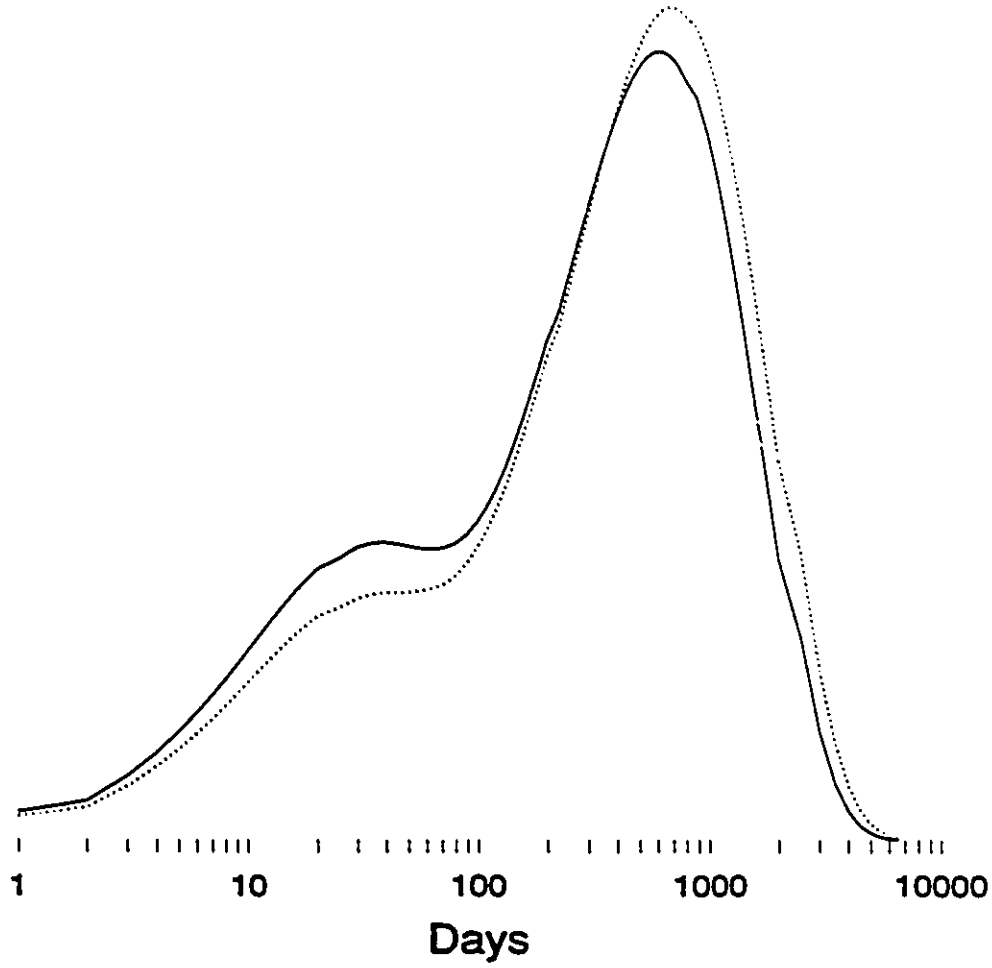

Figure 8. Estimated probability density functions for waiting times between separations - cases and controls

Probability densities based on a 2 compartment model. Solid line is for controls; dotted line is for cases.

External validity of hospital data

The observed and predicted distributions of parity by age group for all controls having a live birth delivery were not significantly different (Table 11). There was an excess of primiparas in the control group, at the expense of para 2's (p=.098). This suggests that, while few historical deliveries are missing from the data set, some first deliveries are missing.

Table 11. Number of controls delivering by age group and parity

	Parity								
Age group	1	2	3	4	5	6	7	8	Total
15-19	15	2	0	0	0	0	0	0	17
20-24	49	31	10	2	0	0	0	0	92
25-29	75	44	27	13	3	1	0	0	163
30-34	36	36	18	7	0	0	1	1	99
35-39	10	8	8	2	3	0	0	1	32
40-44	2	2	1	0	0	0	0	0	5
Total	187	123	64	24	6	1	1	2	408
Predicted	170	148	63	19	5	2	1	0	

Chi-square (5 df) = 7.58, $p > .10$
Chi-square (parity 1 and 2 only, 1 df) = 2.74, $p = .098$
Expected numbers are based on all Canada, 1986 (Appendix D)

Description of ectopic pregnancy episodes

There were 674 valid EP episodes in the three year period April, 1984 through March, 1987. Of the 641 episodes for which admission status was recorded, 85% were admitted on an emergency basis (Table 12). The average hospital stay was 5.4 days. No deaths occurred in hospital; no maternal deaths due to EP were recorded for this period in Saskatchewan. Secondary diagnostic codes were indicative of excessive blood loss for 22 (3.3%) of all EP separations for the three years (Tables 3,5). Of the 616 EP episodes having the time to first procedure coded, 91% had surgery on the first or second hospital day (Table 13). EP episodes led to hospital stays of 5.8 (sd 2.4) days, on average, including transfers between institutions.

Table 12. Admission and separation status on EP and control separations

Admission status	Cases	(%)	Controls	(%)
Emergency	545	81%	367	53%
Urgent	63	9%	110	16%
Elective	33	5%	177	26%
Unknown	33		33	
Separation status	**Cases**	**(%)**	**Controls**	**(%)**
Discharged home	641	95%	663	97%
Left without permission	3	.4%	2	.3%
Transferred	29	4%	21	3%
Unknown	1		1	

Table 13. Days to first procedure for 616 EP episodes with a procedure coded

Operation day	Number of episodes
Day of admission	353
1	204
2	33
≥3	26

Six EP (1%) were coded as ovarian pregnancy (Table 14). Because of the small number and the inability to verify diagnosis they were included in the estimation of odds ratios, even though the risk factors for ovarian pregnancy may differ from tubal pregnancy.

Table 14. Coding of ectopic pregnancy on EP episode separations

Number	%	ICD-9	Diagnosis
6	0.9	633.0	Abdominal pregnancy
548	81.3	633.1	Tubal pregnancy
6	0.9	633.2	Ovarian pregnancy
14	2.1	633.8	Other ectopic pregnancy (cervical, combined, cornual, intraligamentous, mesometric, mural)
100	14.8	633.9	Ectopic pregnancy unspecified

Results of case-control study

Thirty-five individuals (5.5%) had multiple ectopic pregnancies during the 3-year study period (Table 15). There were 634 valid case individuals and 687 control individuals available for analysis. No control could be matched for one episode in the oldest age group and this case was excluded. In the 1986/87 data, two controls were inadvertently selected for each of 11 case episodes having a transfer to another hospital. I arbitrarily excluded the control chosen for the first of the separations, that is, the separation with a transfer discharge code and the earlier separation date. Thus, only the 633 individual pairs or 673 episode pairs were included in the case-control analysis.

Table 15. Ectopic pregnancy episodes per individual, 1984-86

Number of EP	Individuals	Episodes	Controls selected
0	(3 coding errors)*	0	3*
1	1*	1*	0
1	587	587	587
1	11	11	22†
2	31	62	62
3	3	9	9
4	1	4	4
Total	637	674	687
*Total excluded	4	1	14
Total included	633	673	673

* excluded from analysis
† 1 extra control excluded for each of 11 cases

Few characteristics of individuals were available. Those which derive from age or known pregnancy history are presented in Table 16. The age at first known separation was lower and the number of years of known history was greater for cases than for controls, but the number of hospitalizations did not differ. The mean time from first pregnancy and most recent prior pregnancy were also slightly longer for cases. The paired t-test for age was significant in spite of the matching on age. The interpretation of a positive mean difference in age between cases and controls is that, within 5-year age groups, cases tended to be slightly older than controls, which agrees with the increasing rate of EP and declining fertility with age.

Table 16. Characteristics of cases and controls at index separation (episodes)

Characteristic	Cases		Controls		p-value
	Mean	SD	Mean	SD	(paired t)
Age	27.8	5.33	27.7	5.29	0.029
Age at first separation	22.5	6.17	23.2	5.93	0.006
History (years)	5.3	4.91	4.5	4.68	<0.001
History (number)	3.1	4.28	3.2	4.38	0.706
Gravidity	2.2	1.36	2.2	1.48	0.724
Parity excluding index	0.95	1.18	0.99	1.26	0.514
Years since first pregnancy*	5.7	4.21	5.3	3.99	
Years since last pregnancy*	3.1	2.80	3.0	2.69	

* n = 414 cases and 394 controls with Gravidity > 1

On univariate matched analysis of individuals, odds ratios (OR) greater than one were found for the known risk factors for EP (Table 17). Only two case index separations had codes which suggested an IUD was in place at the time of EP; the exact lower 95% confidence limit for the unmatched odds ratio was 0.21. The 95% confidence intervals of OR for endometriosis and induced abortion included 1. Matched OR were also estimated after exclusion of case episodes which did not have a definitive procedure code (Table 18). Univariate matched OR based on EP episodes were comparable to those for individuals, as were the Mantel-Haenszel OR, with stratification by age group (Table 19).

Stratification was used to assess interaction between prior EP and other risk factors (Table 19). Prior EP has often been an exclusion criterion in case-control studies of EP. Antecedents to EP tended to be similar for cases and controls having a prior EP, with odds ratios of 2 or less (Table 20). Confounding by prior EP is seen for tubal repair (OR 21.5 adjusted for EP, compared to 32.8 when adjusted for age only) and other tubal surgery (OR 2.31, compared to 7.08). In both cases, there is also effect modification, although the confidence intervals for the odds ratios (prior EP) are wide.

Adjustment for other factors by multiple logistic regression analysis yielded lower estimates of the odds ratios for chronic PID or pelvic adhesions, prior ectopic pregnancy, tubal sterilization, and tubal repair (Table 21). A logistic model which included other tubal surgery was similar (OR for tubal surgery 1.2 OR for prior EP 7.9, correlation between regression coefficients -0.68).

Etiologic fractions for individual factors ranged from 1.4% to 9.9% (Table 22). The factors represent hospitalization for, or with, a particular condition. For example, 4.7% of EP episodes may be attributed to acute PID *treated in hospital.* Etiologic fractions for *all* PID episodes cannot be estimated, because the true prevalence of (a history of) PID among controls is not known. For

example, assume that ascertainment of PID is as low as 10%. Then the control (population) prevalence of PID could be as high as 29%. Etiologic fractions for a fixed, known, relative risk are very sensitive to the estimated population prevalence (prevalence among controls) as illustrated in Figure 9. In the absence of reasonably precise estimates of either population prevalence or ascertainment, estimates of etiologic fraction are not valid.

The etiologic fraction for hospitalization with any gynaecologic infection (PID or an STD) was 10.1% (approximate 95% CI 3.9 - 16.2), and for any tubal surgery, 17.4% (CI 10.4 - 24.4). The etiologic fraction for any tubal surgery was higher in the 1985/86 and 1986/87 data, but the confidence intervals are wide.

Risk factors are correlated, so that etiologic fractions are not additive. Even though the etiologic fractions are not adjusted for other factors, they are valid for comparing the relative importance of risk factors (which relate to hospitalization)[123].

Table 17. Univariate odds ratios for individuals, matched analysis

Risk factor*	OR†	95% Confidence Interval	
Acute PID	2.88	1.6	5.3
Chronic PID	6.83	2.8	17.8
Gonorrhoea, cervicitis	2.60	0.9	8.3
Endometriosis	0.67	0.3	1.8
Induced abortion	0.94	0.6	1.6
Infertility	3.75	1.9	7.5
IUD in situ (2 cases only)	undefined	--	--
Prior ectopic pregnancy	10.2	4.2	26.1
Sterilization	14.5	3.4	87.9
Sterilization (not reversed)	7.50	2.5	25.1
Tubal repair (any reason)	23.5	5.6	140.
Reversal of tubal ligation	undefined	--	--
Other tubal surgery	6.88	3.2	15.6

* episodes/conditions resulting in hospital stay only
† matched OR for 633 individual pairs, based on most recent EP
-- unable to calculate

Table 18. Univariate odds ratios for episodes with EP procedure coded

Risk factor*	Matched OR†	95% Confidence Interval	
Acute PID	2.79	1.5	5.4
Chronic PID	7.00	2.9	18.3
Gonorrhoea, cervicitis	2.00	0.8	5.5
Endometriosis	0.60	0.1	2.9
Induced abortion	0.88	0.4	1.9
Infertility	3.77	2.0	7.3
Prior ectopic pregnancy	15.0	5.2	48.5
Sterilization	14.0	3.3	85.0
Sterilization (not reversed)	6.50	1.4	41.7
Tubal repair (any reason)	25.5	6.1	151.
Other tubal surgery	11.4	4.4	32.3

* episodes/conditions resulting in hospital stay only
† matched OR based on 573 episode pairs

Table 19. Comparison of matched and stratified odds ratio estimates for episodes

Risk factor*	Matched OR†	M-H OR adjusted for age only	M-H OR adjusted for prior EP and age
Acute PID	2.82	2.75	2.77
Chronic PID	6.57	7.06	5.37
Gonorrhoea, cervicitis	2.29	2.14	1.63
Endometriosis	0.83	0.81	0.63
Induced abortion	0.94	0.95	1.03
Infertility	4.00	4.31	3.15
IUD in situ (2 cases only)	undefined	--	--
Prior ectopic pregnancy	12.0	12.0	--
Sterilization	15.0	15.9	16.6
Sterilization (not reversed)	7.50	7.70	--
Tubal repair (any reason)	28.0	32.8	21.5
Reversal of tubal ligation	undefined	--	--
Other tubal surgery	8.25	7.08	2.31

* episodes/conditions resulting in hospital stay only
† matched OR based on 673 episode pairs
-- unable to calculate

Table 20. Odds ratios adjusted for prior EP and age

Risk factor*	Odds ratio† for prior		OR adjusted for EP and age
	EP=yes§	EP=no	
Acute PID	undefined	2.7	2.8
Chronic PID	1.2	6.6	5.4
Infertility	2.1	3.3	3.2
Sterilization	undefined	15.6	16.6
Tubal repair	1.8	46.2	21.5
Other tubal surgery	1.4	2.6	2.3

* episodes/conditions resulting in hospital stay only
† Mantel-Haenszel Odds Ratio adjusted for 5-year age group
§ 73 cases, 7 controls having prior EP

Table 21. Odds ratios estimated by logistic regression (episodes)

Risk factor*	OR†	95% Confidence Interval§	
Acute PID	2.3	1.1	4.9
Chronic PID	2.7	0.8	9.2
Infertility	2.0	0.8	5.0
Prior ectopic pregnancy	8.9	3.0	26.3
Sterilization	9.4	1.3	69.4
Tubal repair (any reason)	12.4	1.8	87.3

* episodes/conditions resulting in hospital stay only
† adjusted for age group and other factors in model
§ test-based joint Bonferroni confidence interval

Table 22. Etiologic fractions from univariate odds ratios (episodes)

Risk factor*		M-H OR†	Prevalence in controls	EF§
Acute PID		2.8	2.9%	4.7%
Chronic PID		7.1	5.7%	5.9%
Gonorrhoea, cervicitis		2.1	1.2%	1.4%
Any infection				10.1%
Infertility		4.3	2.0%	6.4%
Prior ectopic pregnancy		12.0	1.0%	9.9%
Sterilization		15.9	0.3%	4.2%
Tubal repair (any reason)		32.8	0.3%	8.1%
Other tubal surgery		7.1	1.6%	8.8%
Any tubal surgery	all data			17.4%
	1984 data			16.4%
	1985 data			18.0%
	1986 data			17.9%

* episodes/conditions resulting in hospital stay only
† Mantel-Haenszel Odds Ratio adjusted for 5-year age group
§ etiologic fraction adjusted for age

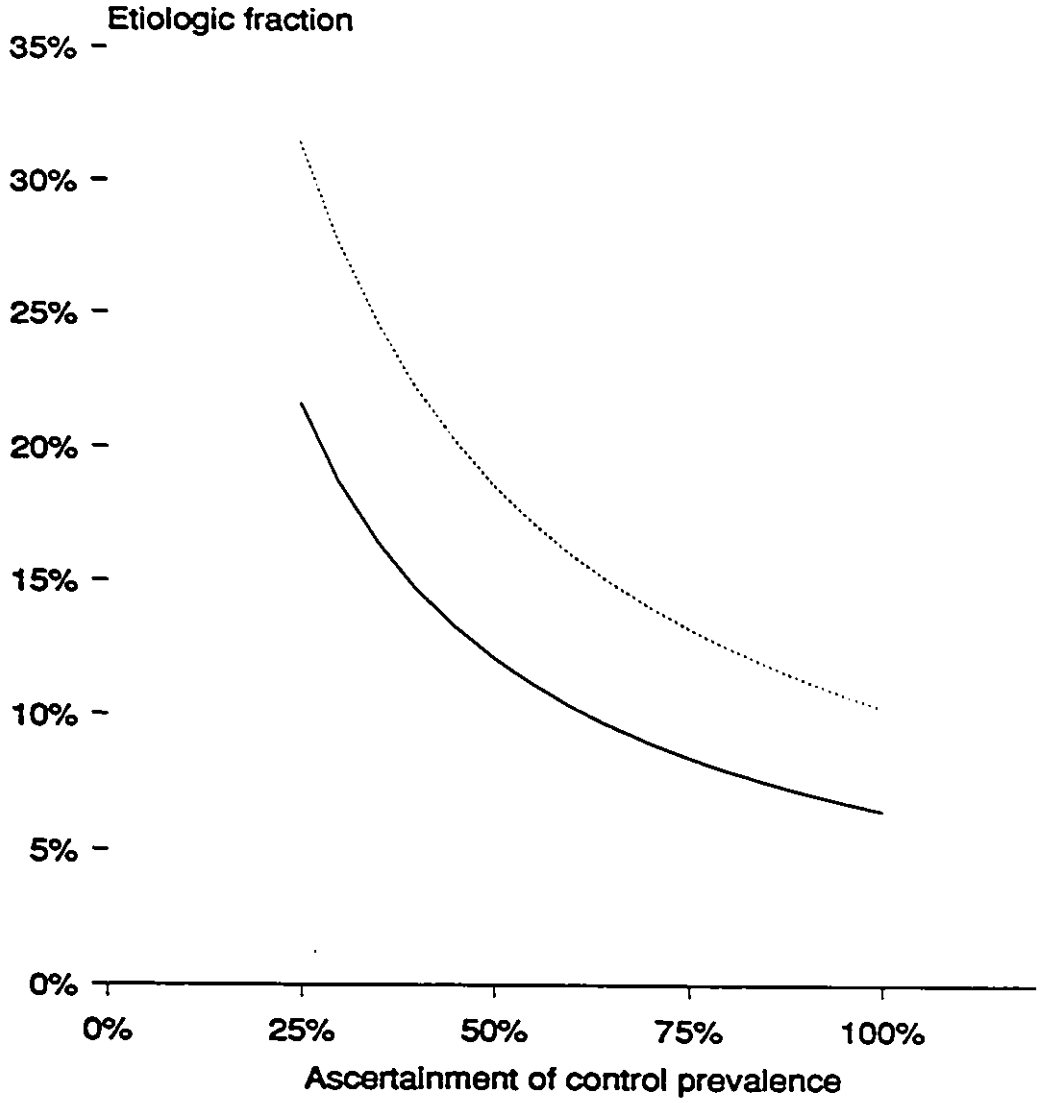

Figure 9. Effect of ascertainment of control exposure on estimate of etiologic fraction

Solid line is infertility, dotted line is prior ectopic pregnancy.

CHAPTER IV

Discussion

Administrative databases have been designed for purposes other than research. The first concern in using such data for research should be the validity of the data. In this case control study, at issue is the accuracy and completeness of recording and coding of diagnostic information in the hospital morbidity database. The gold standard for validity of a clinical diagnosis of tubal ectopic pregnancy would be confirmation of the presence of trophoblastic tissue in a surgical specimen from the fallopian tube. It would be reasonable to require such evidence of EP in a study which had access to hospital charts. In this study, the assumption was made that clinical diagnosis and recording of EP were correct, and if not complete, then at least representative of all EP in Saskatchewan.

One instance of miscoding of EP in hospital morbidity is known. In Prince Edward Island, a small number of live birth deliveries were coded as abdominal pregnancy for a number of years[101]. This coding error was detected only because of the small baseline number of EP per year in Prince Edward Island. An unpublished review of 479 EP separation records for Manitoba (1978-1980/81) found that only 2 (0.4%) had incompatible procedures coded; 38 (7.9%) had no procedure coded; and 421 (87.9%) had a definitive procedure coded (LCDC, unpublished data, 1984). Validation of the SHSP data for 1984-86 led to the identification of only 3 probable errors in coding (Tables 5, 6).

Chart review was not done in this study, but would be of value in assessing the validity of those EP separations without procedure coding

consistent with EP. For index separations having a secondary diagnosis or procedure code, there were very few inconsistencies which would indicate miscoding. Three cases and no controls were excluded on this basis. Fewer than 7% of EP separations had no procedure coded. Re-analysis of the case-control data, limited to cases having a definitive EP procedure, yielded results similar to those obtained with the full data set.

Less is known about the validity of recording and coding of other diagnoses of interest in this study, however. Extensive chart review would be required to validate historical diagnoses for cases and controls. Surgical procedures pose less of a concern than do diagnoses such as spontaneous abortion or PID. Roos suggested that diagnostic codes in hospital morbidity data are valid for severe illness or major surgery[75].

Internal consistency of record linkage

In this study, the record linkage process depended only on the unique client identifier used by Saskatchewan Health. Neither the actual client number nor other identifiers were available to validate the linkage. Errors in date of birth (DOB) occurred in unexpected numbers in the 1986/87 data. These records were extracted close to the time when SHSP considered the data "final". It may be that some errors in transcribing DOB in the current year have since been corrected. This could not explain the occurrence of a much greater proportion of DOB errors found in the historical data linked for cases and controls in that year compared with the previous two years.

Ectopic pregnancy risk factors

For relatively little cost ($4200 for data generation), this case control study was powerful enough to detect significant odds ratios for a number of known risk factors for ectopic pregnancy. Univariate age-adjusted odds ratios

were consistent with those found in previous studies (Table 23). Odds ratios for PID were lower than in the Seattle and Mayo Clinic studies, which used chart review and/or interviews to establish exposure histories. This difference may be explained by under ascertainment of these exposures in the SHSP data, which would, if it occurred equally among cases and controls, bias the odds ratio towards the null[124]. Odds ratios for tubal surgery and sterilization are higher than in other studies, although 95% confidence intervals include the other estimates. Surgical risk factors will be better ascertained in the SHSP data than in studies using chart review at a single hospital or relying on interviews. The matched odds ratios for surgery are also biased upward because of confounding with prior EP. This was most noticeable for "other tubal surgery", for which the odds ratio adjusted for age and prior EP was only 2.31 (episodes), compared to 8.25.

Table 23. Point estimates of relative risk from case-control studies using pregnant hospital controls

Exposure	Study reference			
	SHSP*	WHO[44]	Seattle[45,47,48]	MAYO[51,†]
Past history of PID or STD	3.4	2.8	5.1	4.0
Previous EP	10.2	7.0	excl[§]	excl
Previous induced abortion[‡]	0.9	0.9	1.3-1.8	2.5
Sterilization	14.5	10.9	excl	-
Infertility	3.8	-	2.7	2.9
Tubal surgery	6.9-24	-	-	5.6

* matched odds ratios for individuals
† MAYO - controls selected from clinic and hospital records
§ exclusion criterion for case selection
‡ 95% confidence limits include 1 in all studies listed

The greatest risk for ectopic implantation of a fertilized ovum is in women who have had a prior EP or a tubal repair. High risk was also associated with tubal sterilizations which were not reversed. The latter is of particular importance, since the proportion of Canadian women who use sterilization for contraception increased in the 1970s to over 30%[125].

Biases may limit the validity of case-control studies which use only administrative data. In this study, it was not possible to ascertain exposures which did not result in hospitalization in Saskatchewan. The database captured only two diagnostic codes and one procedure code. Variables which represent "lifestyle" factors, such as smoking, sexual activity, alcohol use, and diet, are rarely included in administrative data. When a missing or poorly determined variable is a confounder, it is not possible to predict the direction of bias.

Ascertainment of exposures

Risk exposures may not be represented in the data, or may be incompletely ascertained. This is the major shortcoming of this study. As was anticipated, low ascertainment limits the interpretation of the etiologic fractions for PID and infection because the measured prevalence of these factors among controls underestimates the population prevalence.

Odds ratios for PID and infections are biased because of this misclassification of exposures. Errors in classification should be almost exclusively among those who have had infections but were never hospitalized. If this under ascertainment is non-differential, then the odds ratio will be biased toward unity. There is some prognostic value to the odds ratio obtained for those women who have been hospitalized for PID and intend to conceive. The odds ratio provides the relative odds of extrauterine versus intrauterine implantation, given fertilization, but does not yield risk of EP, given unprotected sexual activity, because of the choice of controls.

If only those episodes of PID serious, or symptomatic, enough to require hospitalization carry a risk of ectopic pregnancy, then there will be little bias in the estimate obtained in this study. However, current clinical opinion is that even "silent" salpingitis, common with C trachomatis infection, can cause significant tubal damage[108,126].

Ascertainment of prior EP and surgical procedures other than tubal ligation should be complete for those done in Saskatchewan after 1970. A small number of procedures will have been done outside the province or too early to have been recorded in the data, but there is no reason to believe this would occur differently for cases or controls. There should be little misclassification bias, then, in the odds ratios for prior EP and tubal surgery. Furthermore, the study results can be used to estimate etiologic fraction for these factors. This estimation is useful from the perspective of surveillance and understanding recent trends. The data suggest that over 17% of EP in Saskatchewan in 1984-86 may be attributed to various forms of tubal surgery (Table 22). It was not possible, in the three years studied, to detect more than a suggestion of an increasing trend in this proportion.

Risk factors not represented in the data include smoking, contraceptive use and congenital anomalies of fallopian tubes. The etiologic fraction of cases due to these factors cannot be estimated from this study.

Studies need not be limited by the absence of data on effect modifiers or other independent risk factors. This may be the case with smoking and ectopic pregnancy. If so, then valid estimates of relative risk may be obtained in the absence of data on smoking history.

Confounders

Confounding of some variables with prior EP was apparent in the data, as mentioned above. Adjustment for confounding with prior EP required the use of stratified analysis or logistic regression.

If unmeasured variables are confounders of the relationship between EP and a factor X, it is not possible to predict the direction of the bias in the estimation of risks due to X. Potential confounders of the relationship between EP and PID which are not in the SHSP database include smoking and current IUD use. Some authors have suggested that smoking might exert its risk via impaired immunity and salpingitis[49], that is:

smoking --> impaired immunity --> PID --> EP

If smoking and PID share a single causal pathway as illustrated, then adjustment of PID risk for smoking would be inappropriate.

The ICD provides for coding of an IUD in situ, yet only two such codes were found for index ectopic pregnancies, and none for control index pregnancies. Current IUD use may be a confounder because of its association with PID and EP. The positive association with EP is conditional on fertilization; IUD use protects against all forms of pregnancy, conditional on sexual activity. The reported prevalence of current IUD use among cases of EP has varied greatly, depending on the population studied. No reliable historical estimates of IUD use are available for Canada[127,125].

A possible confounder of EP and infertility is the use of fertility drugs. Only women with infertility will be given drugs such as clomiphene citrate. In the Mayo Clinic study, clomiphene use was confounded with infertility, but it was possible to examine the effect of clomiphene within the group of infertile women. A small increased risk associated with the drug was detectable[51]. It was not possible to assess the contribution of infertility drugs to the odds ratio in the SHSP study, but this could be done if hospital data were linked to the drug database in Saskatchewan.

One remedy to the absence of confounders is to restrict the population being studied. For example, in a study of adverse pregnancy outcomes in relation to drug use, all cases would clearly have been sexually active. Were that study to sample the entire female population for controls, the relationship between drug use and outcome would be confounded by sexual activity, provided drug use was related to sexual activity. Another example would be a study of an infertility drug and congenital anomalies. Use of fertility drugs is associated with pregnancy, otherwise there would be no indication for their use. Among the general population of women, bearing a child with a congenital anomaly is clearly associated with pregnancy. In this example, restricting the case and referent population to pregnant women will avoid confounding by conception.

Other sources of bias

Migration may lead to information bias, but it is difficult to see how there could have been differential migration between cases and controls in this study. Interprovincial migration is related to age, but this study was stratified on age.

There may be selection bias if the referent population is not completely represented in the data. In this study, women having induced abortion were in the referent population of conceiving women. Very few were eligible to be selected as controls, since most induced abortions are done as day procedures. On the other hand, almost all women having a live birth were eligible for control selection. Bias may occur if the prevalence of a risk factor is different for women having induced abortions than for other pregnant women.

Most studies of EP have been unable to demonstrate an association between EP and induced abortion. It does not follow that induced abortion is unrelated to risk factors for EP, however. Failure to included women having induced abortion among controls will lead to inflated odds ratios for factors with

a positive correlation with induced abortion. This may be true for sterilization. A woman having a tubal sterilization may be less likely to continue a subsequent pregnancy. The situation may be reversed for other forms of tubal surgery. A woman having reconstructive tubal surgery may be more likely to wish to carry a subsequent pregnancy to term. This study would then bias the odds ratio for tubal repair towards unity.

For those factors which may also lead to infertility, the choice of pregnant or non-pregnant controls will depend on the inferences one wishes to make (see Chapter I, page 18).

Estimates of relative risk (odds ratios) obtained from the logistic regression are subject to bias if the logistic model is not appropriate[128]. The appropriate goodness-of-fit statistic (Hosmer-Lemeshow) for the model shown in Table 21 was not significant, suggesting that the observed data fit a logistic model. The model accounts for a small proportion of the variation because of the low ascertainment of exposures. However, the odds ratios obtained are reasonable, given those reported in the literature.

Ectopic pregnancy surveillance using administrative data

Readily available administrative and vital statistical data currently provide reliable estimates of EP incidence for all of Canada. Similar estimates can be obtained for each province with little difficulty.

The trend to increasing incidence is similar to that found throughout the developed world. The trend is similar for all provinces, based on published hospital morbidity data. The most recent surveillance data from the United States shows that the incidence of EP may have levelled off in 1986[17]. The only factor for which surveillance data available in Canada, and which could have contributed to the increasing incidence of EP is maternal age. Increased age carries with it an increased risk of EP and the trend towards delayed

childbearing in Canada could contribute to the EP trend. However, as already shown, the increasing trend is present in all age groups, and age-standardized rates show that at most 13.6% of the increase between 1971 and 1985 could be due to changing maternal age. A similar finding has been reported in Sweden[26].

Surveillance of EP in the United States is based on data from the National Hospital Discharge Survey which collects details on all separations from a national sample of hospitals[17]. This survey provides more detailed data than are available in the hospital separation databases in Canada. Incidence estimates for the United States are available by age, race, and region.

Existing hospital morbidity data for Canada overestimate the number of EP episodes because transfers may result in multiple counting of a single episode and suspected, but never proven, diagnoses may be included. For the year 1984/85, the Statistics Canada hospital morbidity summary lists 250 EP for the province of Saskatchewan. With the elimination of transfers from the SHSP data set, there were actually 231 EP episodes in that year. The routine surveillance data had an overcount of 8.2%.

On the other hand, hospital morbidity data may become less complete for EP in the future. Clinicians now recognize that not all EP require surgery or a hospital stay[129,130]. Hospital morbidity data may greatly underestimate the incidence of EP when chemotherapy or laparoscopic techniques for removal of uncomplicated EP are done as day procedures. Alternative data sources must be available in order to monitor such changes in clinical practice in order to validate trends derived from hospital morbidity data.

Utility of hospital data for etiological research

Hospital morbidity data such as those available from HMRI can be used to study relationships among factors which are uniquely represented by ICD

diagnostic coding on separation records. Confounders of the relationship of interest must also be found in the data in order to study the relationship. The database must cover enough years to include exposures and confounders. The existing HMRI database is limited because encrypted identifying numbers may not be unique. Halliday's algorithm is only applicable to studies of utilization for a single condition, or to estimation of prevalence from hospital morbidity records[65]. It could be extended to provide partial histories on an individual by varying the sorting specifications, but only for one variation of identifier in provinces, such as Ontario, where these are not unique. Furthermore, the encryption procedure results in a zero ID code for some individuals, presumably when the actual identifier is missing. Changes in identifier for women whose marital status changes or for persons reaching the age of majority cannot be tracked in order to complete histories. This tracking can only be done by provincial health insurance or hospital services commissions. It is not possible to track individuals across provincial boundaries.

Individual histories compiled by record linkage within a province are censored prior to the time during which a person was covered by the provincial hospital care plan. From the SHSP data alone, the length of coverage by the plan cannot be determined. The earliest known hospital separation may represent the first hospitalization, or just the first in Saskatchewan. This study design assumes that no censoring occurs, so that a missing exposure is treated as no exposure. The provincial registry could be used to determine the dates of coverage in order to assess censoring.

The HMRI hospital morbidity database for 1978 through 1987 is now available at the Laboratory Centre for Disease Control in a format which can be

used on a microcomputer*. Briefly, certain fields, including the encrypted identifier, were extracted from all records and transferred to a compact disk (CD-ROM) format. The files are indexed by diagnosis, province, age group, sex and year. Extraction of 2100 records for ectopic pregnancies in the province of Ontario in 1987 took approximately 150 minutes on an IBM-compatible 80386 computer. The limiting factor is the slow transfer rate of data from the CD-ROM. For personal computer use, additional indexing (identifier, date of birth) and the use of a packed data format would make data retrieval more efficient[131].

The present study was done to explore the use of hospital morbidity data as a surveillance tool. The record linkage was limited to prior hospital separations in order to reduce the cost. With low data acquisition costs, the study could be repeated, perhaps on an annual basis. If the results are used to analyze a trend, some bias may be acceptable, provided one can expect the bias to be consistent over time. Record linkage to other databases can provide a much richer data set, but at greater cost. Linkage between databases has been done only by provincial health insurance commissions in jurisdictions with unique identifiers. A much broader spectrum of studies is possible, as reviewed earlier, but may be too expensive to repeat over time as a surveillance tool. For example, the SHSP unique identifier could have been used to link to historical records of physician billing diagnoses. As was pointed out earlier, diagnoses provided on physician billing records may be imprecise. Validation of physician billing diagnoses for cases and controls would be difficult and very expensive.

Record linkage to the physician billing database for control selection would provide a means of identifying women having a pregnancy outcome

---

* Further information on the Hospital Medical Records Institute database at LCDC is available from the author, or from the Director, Bureau of Chronic Diseases, LCDC, Tunney's Pasture, Ottawa, Ontario, Canada, K1A 0L2

(induced or spontaneous abortion) which did not result in a hospital stay. The hospital and physician billing records could be used to create a list of all women having any pregnancy-related diagnosis or procedure. Controls would then be selected from this list.

Personal contact with cases poses an ethical problem and may only be justified if done prior to accessing the database[91]. However, one study using a Canadian database did involve contacting patients identified from an administrative database[90]. There is not the same problem with confidentiality in a study which uses the database for follow up of known individuals. This type of follow-up study has application in situations where cohort membership is determined by some event which is not represented in a linkable database. An example would be follow up of immunized children for serious adverse reactions or the occurrence of disease which should be prevented by immunization.

Using record linkage methodology, all provincial hospital databases have the potential to provide individual histories of hospitalization within the province. The literature cited in Chapter I included studies in Newfoundland, Quebec, Ontario, Manitoba, Saskatchewan and British Columbia. The existence of a truly unique identifier makes this simpler. In Manitoba and Saskatchewan, the provincial hospital services plans have demonstrated a willingness to facilitate such research. Simple linkage of records is adequate for some study designs, but a complete population registry will allow determination of person-time denominators in follow up studies and an indication of censoring in retrospective studies. A population registry is also required to track individual data across changes in personal identifier.

CHAPTER V

Conclusions

The case-control study design using only administrative data has been used for post-marketing surveillance of drugs, but has had few other applications because of limited ascertainment of exposures. The obvious advantages of providing large amounts of data (power), at low cost, on whole populations (external validity) must be weighed against the limitations imposed by the potential lack of confounders and information bias in the databases available. In Canada, there is a positive outlook on the use of administrative and other databases containing personal identifiers for research, and safeguards have been implemented to prevent the release of confidential information. Record linkage capabilities are built into many databases, in part because of the nature of the identifying data, but also by design.

Hospital morbidity databases have been the major source of passive, aggregate, surveillance data on ectopic pregnancy in Canada. These data have shown that the incidence of ectopic pregnancy, as measured by separations, has increased in this country in the past two decades to the same extent as in most other countries where surveillance data have been reported. Internal linking of hospital separation is required to identify episodes and provide an incidence estimate. Additional surveillance for outpatient treatment of ectopic pregnancy must be instituted to validate future trends based on hospital morbidity data.

A number of cohort and case-control studies have established the risk of ectopic pregnancy for a number of known risk factors, but none has been able

to account for changes in trends for risk factors for ectopic pregnancy. There is no report in the literature of a case-control study of ectopic pregnancy which used administrative data exclusively. The feasibility of such a study has now been demonstrated. Additional diagnostic and procedure codes were used to validate case and control selection. The potential biases in the estimation of relative risk can considered acceptable, but not without reservation. This applies to the odds ratios for risk factors which may be poorly represented in the data, namely tubal infections and infertility. However, the estimates obtained were consistent with those from other case-control studies which used similar (pregnant) controls. In general, such estimates should be regarded with suspicion and confirmed by additional data sources. Population attributable risk cannot be reliably estimated for poorly ascertained factors.

Because of the low cost and large population base, the SHSP data provide a useful surveillance tool with which to observe changes in the trends of ectopic pregnancy and some risk factors for ectopic pregnancy. These data derive from a well-defined population and can be used to estimate population attributable risk proportions for some risk factors, namely, previous ectopic pregnancy and tubal surgery. The data have been subjected to validity checks, so that extraction of exposure data may be automated, with limited intervention required on the part of the researcher.

The SHSP data set can be a useful surveillance tool, particularly for measuring trends in ectopic pregnancy following tubal surgery. The data cannot be used to follow trends in the etiologic fraction of ectopic pregnancy due to pelvic inflammatory disease, however. Thus, the case-control design cannot provide an explanation for all of the increasing trend in ectopic pregnancy without additional data sources. Data for more years are required to determine the contribution of tubal surgery to the increasing trend in ectopic pregnancies. The same study may be done in Manitoba now, and in other provinces once

sufficient data have been acquired in an organized fashion to permit construction of longitudinal histories on individuals.

This study demonstrates the high risk of EP associated with tubal damage, from infection, prior ectopic pregnancy, and destructive or reconstructive surgery. Women known to have tubal damage from any cause should know the risk of ectopic pregnancy associated with these conditions and be aware of the symptoms of early pregnancy so they can receive adequate early prenatal assessment.

References

1. Hockberger RS. Ectopic pregnancy. Emergency Clinics of North America 1987;5(3):481-493.

2. Droegemuller W. Ectopic pregnancy. In: Danforth DN, Scott JR, eds. Obstetrics & Gynecology. 5th Ed. Philadelphia: JB Lippincott, 1986.

3. Barnes AB, Wennberg CN, Barnes BA. Ectopic pregnancy: incidence and review of determinant factors. [Review] Obstet Gynaecol Surv 1983; 38(6):345-56.

4. Hockin JC, Jessamine AG. Trends in ectopic pregnancy in Canada. Can Med Assoc J 1984;131(7):737-40.

5. Laboratory Centre for Disease Control. Sexually Transmitted Disease in Canada 1986. Can Dis Weekly Rep 1988;14(Supp 1):1S.

6. Chavkin W. The rise in ectopic pregnancy--exploration of possible reasons. Int J Gynaecol Obstet 1982;20(4):341-50.

7. Chow WH, Daling JR, Cates W Jr, Greenberg RS. Epidemiology of ectopic pregnancy. Epidemiol Rev 1987;9:70-94.

8. Kim DS, Chung SR, Park MI, Kim YP. Comparative review of diagnostic accuracy in tubal pregnancy: a 14-year survey of 1040 cases. Obstet Gynecol 1987;70(4):547-54.

9. Andolsek KM. Ectopic pregnancy: 'classic' vs common presentation. J Fam Pract 1987;24(5):481-5.

10. Gorodeski IG, Bahary CM. Tubal pregnancy--reappraisal of incidence. Euro J Obstet Gynecol Repro Biol 1987;24(1):57-62.

11. Schuman EA. Extrauterine Pregnancy. New York: D Appleton & Co., 1921. (reference from Barnes et al)

12. Dorfman SF. Ectopic pregnancy surveillance. MMWR - Surveillance Summaries 1983;32(1):19SS-21SS.

13. MacKay HT, Hughes JM, Hogue CR. Ectopic pregnancy in the United States, 1979-1980. MMWR CDC Surveillance Summaries 1984;33(2):1SS-7SS.

14. Atrash HK, Hughes JM, Hogue CJ. Ectopic pregnancy in the United States, 1970-1983. MMWR CDC Surveillance Summaries 1986;35(2):29SS-37SS.

15. Ellerbrock T, Atrash HK, Hogue CJ, Hughes JM. Ectopic pregnancy mortality in the United States, 1979-1982. MMWR CDC Surveillance Summaries 1987;36(2):13-8.

16. Lawson HW, Atrash HK, Saftlas AF, Franks AL, Finch EL, Hughes JM. Ectopic pregnancy surveillance, United States, 1970-1985. MMWR CDC surveillance summaries, 1988; 37(SS-5):9-18.

17. Lawson HW, Atrash HK, Saftlas AF, Finch EL. Ectopic pregnancy in the United States 1970-1986. MMWR CDC Surveillance Summaries 1989; 38(SS-2):1-10.

18. Rubin GL, Peterson HB, Dorfman SF, Layde PM, Maze JM, Ory HW, Cates W Jr. Ectopic pregnancy in the United States 1970 through 1978. JAMA 1983;249(13):1725-9.

19. Gonzalez FA, Waxman M. Ectopic pregnancy. A retrospective study of 501 consecutive patients. Diagnostic Gynecol Obstet 1981;3(3):181-6.

20. Budnick LD, Pakter J. Ectopic pregnancy in New York City, 1975-1980. Am J Pub Health 1982;72(6):580-4.

21. Shiono PH, Harlap S, Pellegrin F. Ectopic pregnancies: rising incidence rates in Northern California. Am J Pub Health 1982;72(2):173-5.

22. Glebatis DM, Janerich DT. Ectopic pregnancies in upstate New York. JAMA 1983;249(13):1730-5.

23. Hemminki E, Heinonen PK. Time trends of ectopic pregnancies. Br J Obstet Gynaecol 1987;94(4):322-7.

24. Mäkinen JI. Ectopic pregnancy in Finland 1967-83: a massive increase. Br Med J 1987;294(6574):740-1.

25. Mäkinen JI. Increase of ectopic pregnancy in Finland--combination of time and cohort effects. Obstet Gynecol 1989;73(1):21-4.

26. Meirik O. Ectopic pregnancy during 1961-78 in Uppsala county, Sweden. Impact of demographic factors on overall incidence. Acta Obstet Gynecol Scand 1981;60(6):545-8.

27. Andreasen EE. Births, abortions and extrauterine pregnancies in the community of Funen during 1974-1986 [Dan] Ugeskrift For Laeger 1989;151(6):384-7.

28. Herbertsson G, Magnusson SS, Benediktsdottir K. Ovarian pregnancy and IUCD use in a defined complete population. Acta Obstet Gynecol Scand 1987;66(7):607-10.

29. Flett GM, Urquhart DR, Fraser C, Terry PB, Fleming JC. Ectopic pregnancy in Aberdeen 1950-1985. Br J Obstet Gynaecol 1988;95(8):740-6.

30. Clark K, Baranyai J. Pelvic infection and the pathogenesis of tubal ectopic pregnancy. Austral New Zealand J Obstet Gynaecol 1987;27(1):57-60.

31. Macintosh MC. Trends in ectopic pregnancy in New Zealand. Austral New Zealand J Obstet Gynaecol 1986;26(2):145-8.

32. Siskind V. Ectopic pregnancy in Queensland, 1971-1980. Med J Austral 1985;142(13):673-4.

33. Marsden DE, Correy JF. A review of ectopic pregnancies in Southern Tasmania. Austral New Zealand J Obstet Gynaecol 1980;20(4):214-8.

34. Parazzini F, La Vecchia C, Fasoli M, Cecchetti G, Mezzanotte G. Trends in ectopic pregnancies and use of intrauterine devices in Lombardy, Italy 1979-1983. Contraception 1988;37(1):29-38.

35. Szymanski W, Gustowski A, Grabiec M. Clinico-statistical analysis of ectopic pregnancies [Pol] Wiadomosci Lekarskie 1982;35(6):395-400.

36. Gorodeski IG, Bahary CM. Tubal pregnancy--reappraisal of incidence. Euro J Obstet Gynecol Repro Biol 1987;24(1):57-62.

37. Zhang ZM, Qiu SH, Weng LJ, Jing XP. An epidemiological study on ectopic pregnancy associated with IUD. A matched case-control study. Chinese M J - Peking 1988;101(2):143-7.

38. Egwuatu VE, Ozumba BC. Unexpectedly low ratio and falling incidence rate of ectopic pregnancy in Enugu, Nigeria, 1978-1981. Int J Fertil 1987;32(2):113-5, 119-21.

39. Corson SL, Batzer FR. Ectopic pregnancy. A review of the etiologic factors. J Repro Med 1986;31(2):78-85.

40. Weinstein L, Morris MB, Dotters D, Christian CD. Ectopic pregnancy--a new surgical epidemic. Obstet Gynecol 1983;61(6):698-701.

41. Tuomivaara L, Kauppila A, Puolakka J. Ectopic pregnancy--an analysis of the etiology, diagnosis and treatment in 552 cases. Arch Gynecol 1986;237(3):135-47.

42. Weiss NS, Daling JR, Chow WH. Control definition in case-control studies of ectopic pregnancy. Am J Pub Health 1985;75(1):67-8.

43. Edelman DA, Porter CW. The intrauterine device and ectopic pregnancy. Adv Contraception 1986;2(1):55-63.

44. The World Health Organization's Special Programme of Research, Development and Research Training in Human Reproduction: Task Force on Intrauterine Devices for Fertility Regulation. A multinational case-control study of ectopic pregnancy. Clin Repro Fertil 1985; 3(2):131-43.

45. Daling JR, Chow WH, Weiss NS, Metch BJ, Soderstrom R. Ectopic pregnancy in relation to previous induced abortion. JAMA 1985;253(7):1005-8.

46. Chow WH, Daling JR, Weiss NS, Moore DE, Soderstrom RM, Metch BJ. IUD use and subsequent tubal ectopic pregnancy. Am J Pub Health 1986;76(5):536-9.

47. Yang CP, Chow WH, Daling JR, Weiss NS, Moore DE. Does prior infertility increase the risk of tubal pregnancy?. Fertil Steril 1987;48(1):62-6.

48. Sherman KJ, Chow WH, Daling JR, Weiss NS. Sexually transmitted diseases and the risk of tubal pregnancy. J Repro Med 1988;33(1):30-4.

49. Chow WH, Daling JR, Weiss NS, Voigt LF. Maternal cigarette smoking and tubal pregnancy. Obstet Gynecol 1988;71(2):167-70.

50. Mueller BA, Daling JR, Weiss NS, Moore DE, Spadoni LR, Soderstrom RM. Tubal pregnancy and the risk of subsequent infertility. Obstet Gynecol 1987;69(5):722-6.

51. Marchbanks PA, Annegers JF, Coulam CB, Strathy JH, Kurland LT. Risk factors for ectopic pregnancy. A population-based study. JAMA 1988;259(12):1823-7.

52. Ory HW. Ectopic pregnancy and intrauterine contraceptive devices: new perspectives. The Women's Health Study. Obstet Gynecol 1981;57(2):137-44.

53. Burkman RT, Mason KJ, Gold EB. Ectopic pregnancy and prior induced abortion. Contraception 1988;37(1):21-7.

54. Thorburn J, Berntsson C, Philipson M, Lindblom B. Background factors of ectopic pregnancy. I. Frequency distribution in a case-control study. Euro J Obstet Gynecol Repro Biol 1986;3(5-6):321-331.

55. Thorburn J, Philipson M, Lindblom B. Background factors of ectopic pregnancy. II. Risk estimation by means of a logistic model. Euro J Obstet Gynecol Repro Biol 1986;3(5-6):333-40.

56. Schlesselman JJ. Case-Control Studies. New York: Oxford University Press, 1982:236.

57. Handler A, Davis F, Ferre C, Yeko T. The relationship of smoking and ectopic pregnancy. Am J Pub Health 1989;79(9):1239-1242.

58. Weström L, Bengtsson LP, Mårdh PA. Incidence, trends, and risks of ectopic pregnancy in a population of women. Br Med J 1981;282(6257):15-8.

59. Beral V. An epidemiological study of recent trends in ectopic pregnancy. Br J Obstet Gynaecol 1975;82:775-782.

60. The Use of Administrative Social Data: Overview and Guide. Ottawa: Statistics Canada, 1985:B1.

61. Statistics Canada. Hospital Morbidity 1983-84 and 1984-85. Ottawa: Supply and Services Canada, 1989. (Cat no. 82-206)

62. Wigle DT. Selected health status indicators. Chronic Disease in Canada 1986;7(1):5-6.

63. Wilkins K, Morris S, Lane R. Mortality and morbidity of Canada's elderly population: a historical perspective. Chronic Disease in Canada 1988;9(5):79-84.

64. Kreiger N. Osteoporosis in an aging population. Chronic Disease in Canada 1988;9(5):85-87.

65. Halliday ML, Corey PNJ, Coates RA, Rankin JG. A method for estimating "persons" versus "cases" from hospital morbidity data in the absence of unique personal identifiers. Am J Epidemiol 1987;125(5):885-891.

66. Tilson HH. Getting down to bases: record linkage in Saskatchewan [Editorial]. Can J Pub Health 1985;76:222-223.

67. Anderson GM, Lomas J. Determinants of the increasing cesarean birth rate: Ontario data 1979 to 1982. N Engl J Med 1984;311:887-892.

68. Anderson GM, Lomas J. Recent trends in cesarean section rates in Ontario. Can Med Assoc J 1989;141:1049-1053.

69. Anderson GM, Lomas J. Monitoring the diffusion of a technology: coronary artery bypass surgery in Ontario. Am J Pub Health 1988;78:251-254.

70. Anderson GM, Lomas J. Regionalization of coronary artery bypass surgery: effects on access. Med Care 1989;27(3):288-296.

71. Smith M, Newcombe HB. Automated follow-up facilities in Canada for monitoring delayed health effects. Am J Pub Health 1980;70(12):1261-1268.

72. Smith ME. Studies relating to uses of the mortality data base file. In: Howe GR, Spasoff RA, eds. Proceedings of the Workshop on Computerized Record Linkage in Health Research. Toronto: University of Toronto Press, 1986:249-257. (a brief summary of 51 studies using the Canadian Mortality Data Base)

73. Howe GR, Lindsay J. A generalized iterative record linkage computer system for use in medical follow-up studies. Comput Biomed Res 1981;14:327-40.

74. Wajda A, Roos LL. Simplifying record linkage: software and strategy. Comput Biol Med 1987;17:239-248.

75. Roos LL, Roos NP, Cageorge SM, Nicol JP. How good are the data? Reliability of one health care data bank. Med Care 1982;20(3):266-276.

76. Roos LL, Roos NP. Assessing existing technologies: The Manitoba study of common surgical procedures. Med Care 1983;21(4):454-462.

77. Roos LL, Nicol JP. Building individual histories with registries: A case study. Med Care 1983;21(10):955-969.

78. Segovia J, Bartlett RF, Edwards AC. Using data linkage to assess the associations between health status, health practices and medical care utilization. In: Carpenter M, Fair ME, eds. Canadian Epidemiology Research Conference - 1989. Proceedings of the record linkage sessions & workshop. Ottawa: Statistics Canada, 1990:119-124.

79. Studney DR, Hakstian AR. A comparison of medical record with billing diagnostic information associated with ambulatory medical care. Am J Pub Health 1981;71(2):145-149.

80. Wenneberg JE, Roos NP, Sola L, Schori A, Jaffe R. Use of claims data systems to evaluate health care outcomes: mortality and reoperation following prostatectomy. JAMA 1987;257:933-936.

81. Shapiro S. The epidemiological evaluation of drugs. Acta Medica Scand Suppl 1984;685:23-27.

82. Strand LM. Drug epidemiology resources and studies: the Saskatchewan database. Drug Information J 1985;19:253-256.

83. Skegg DCG. Relation between drug utilization and morbidity: A record linkage study. Acta Medica Scand Suppl 1984;683:23-27.

84. Strom BL, Morse ML. Use of computerized databases to survey drug utilization in relation to diagnoses. Acta Medica Scand Suppl 1988;721:13-20.

85. Stergachis AS. Record linkage studies for postmarketing drug surveillance: data quality and validity considerations. Drug Intell Clin Pharm 1988;22(2):157-160.

86. Guess HA, West R, Strand LM, Helston D, Lydick EG, Bergman U, Wolski K. Fatal upper gastrointestinal hemorrhage or perforation among users and nonusers of nonsteroidal anti-inflammatory drugs in Saskatchewan, Canada 1983. J Clin Epidemiol 1988;41(1):35-45.

87. Ross RK, Paganini-Hill A, Gerkins VR, Mack TM, Pfeffer R, Arthur M, Henderson BE. A case-control study of menopausal estrogen therapy and breast cancer. JAMA 1980;243:1635-1639.

88. Findle WD. Studies of drug effects within the Kaiser Foundation Health Plan: Southern California region. Drug Information J 1985;19:243-247.

89. Monson RR. Occupational hazards and fetal deaths. In: Porter IH, Hook EP, eds. Human Embryonic and Fetal Death. New York: Academic Press, 1980:159-164.

90. Hogan DJ, Strand LM, Lane PR. Isotretinoin therapy for acne: a population-based study. Can Med Assoc J 1988;138:47-50.

91. West R, Sherman GJ, Downey W. A record linkage study of valproate and malformations in Saskatchewan. Can J Pub Health 1985;76:226-228.

92. Johnson K, Sherman G. A study of anticonvulsant drug therapy during pregnancy and adverse reproductive outcome utilizing two Saskatchewan health databases (Abstract). In: Howe GR, Spasoff RA, eds. Proceedings of the Workshop on Computerized Record Linkage in Health Research. Toronto: University of Toronto Press, 1986:240.

93. Goldacre MJ, Holford TR, Vessey MP. Cardiovascular disease and vasectomy. Findings from two epidemiologic studies. New Engl J Med 1983;308(14):805-808.

94. Ray WA, Griffin MR, Downey W, Melton LJ. Long-term use of thiazide diuretics and risk of hop fracture. Lancet 1989;i:687-690.

95. Cohen MM, Hammarstrand KM. Papanicolou test coverage without a cytology registry. Am J Epidemiol 1989;129(2):388-394.

96. Cohen MM. Using claims data for case-control study. In: Carpenter M, Fair ME, eds. Canadian Epidemiology Research Conference - 1989. Proceedings of the record linkage sessions & workshops. Ottawa: Statistics Canada, 1990:29-35.

97. L'Abbé KA, Hoey JR, Hanley J, Wacholder S, Nantel A. check title. Am J Pub Health 1988;78:1489-1491.

98. National Center for Health Statistics. International classification of diseases, adapted for use in the United States, eight revision. Washington, DC: USDHEW, PHS, 1968 (PHS publication no 1693).

99. International Classification of Diseases, 1975 Revision. Geneva: World Health Organization, 1978.

100. Statistics Canada. Canadian Classification of Diagnostic, Therapeutic, and Surgical Procedures. Ottawa: Supply and Services Canada, 1980 (cat no 82-562E).

101. Laboratory Centre for Disease Control. Ectopics on the increase in Prince Edward Island? Canada Diseases Weekly Report 1982;8:101-102.

102. Ross SM. Stochastic Processes. New York: Wiley, 1983:23.

103. BMDP3R Non-linear regression. In: Dixon WJ, et al, eds. BMDP Statistical Software. Berkley: University of California Press, 1983.

104. Statistics Canada. Vital Statistics, vol 1: Births. Ottawa: Dept of Supply and Services, 1986. (cat no 84-204).

105. Fleiss JL. Statistical Methods for Rates and Proportions (2 ed). New York: John Wiley & Sons, 1981:42.

106. Brunham RC, Binns B, McDowell J, Paraskevas M. Chlamydia trachomatis infection in women with ectopic pregnancy. Obstet Gynecol 1986;67:722-6.

107. Cates W, Holmes KK. Sexually transmitted diseases. In: Last J, ed. Maxcy-Rosenau Public Health and Preventive Medicine 12th ed. Norwalk, Conn: Appleton-Century-Crofts; 1986:269-70.

108. Svensonn L, Mårdh P-A, Ahlgren M, Nordenskjöld F. Ectopic pregnancy and antibody to Chlamydia trachomatis. Fertil Steril 1985;44:313-7.

109. Estany A, Todd M, Vasquez M, McLaren R. Early detection of genital chlamydial infection in women: an economic evaluation. Sexually Trans Dis 1989;16:21-27.

110. Washington AE, Arno PS, Brooks MA. The economic cost of pelvic inflammatory disease. JAMA 1986;255:1735-1738.

111. Dean AD, Dean JA, Burton AH, Dicker RC. Epi Info, Version 5: a word processing, database, and statistics program for epidemiology on microcomputers. Stone Mountain, Georgia: USD, Incorporated, 1990.

112. SPSS/PC+: SPSS for the IBM PC/XT. Chicago: SPSS Incorporated, 1984.

113. Schlesselman JJ. Case-Control Studies. New York: Oxford University Press, 1982:272.

114. Schlesselman JJ. Case-Control Studies. New York: Oxford University Press, 1982:112.

115. Fleiss JL. Statistical Methods for Rates and Proportions (2 ed). New York: John Wiley & Sons, 1981:116-7.

116. Robins J, Greenland S, Breslow NE. A general estimator for the variance of the Mantel-Haenszel odds ratio. Am J Epidemiol 1986;124:719-23.

117. Mehta CR, Patel NR, Gray R. Computing an exact confidence interval for the common odds ratio in several 2x2 contingency tables. J Am Stat Assoc 1985;80:969-973.

118. BMDPLR Stepwise logistic regression. In: Dixon WJ, et al, eds. BMDP Statistical Software. Berkley: University of California Press, 1983.

119. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic Research. New York: Van Nostrand Reinhold, 1982:440.

120. Neter J, Wasserman W, Kutner MH. Applied Linear Regression Models. Boston: Irwin, 1989:602.

121. Taylor JW. Simple estimation of population attributable risk from case-control studies. Am J Epidemiol 1977;106:260.

122. Schlesselman JJ. Case-Control Studies. New York: Oxford University Press, 1982:224-6.

123. Walter SD. Prevention for multifactorial diseases. Am J Epidemiol 1980;112:409-416.

124. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic Research. New York: Van Nostrand Reinhold, 1982:230.

125. Balakrishnan TR, Krotki K, Lapierre-Adamcyk E. Contraceptive use in Canada, 1984. Fam Plann Perspect 1985;17:209-15.

126. Cumming DC, Honoré LH, Scott JZ, Williams KE. Microscopic evidence of silent inflammation in grossly normal fallopian tubes with ectopic pregnancy. Int J Fertil 1988;33:324-8.

127. Balakrishnan TR. Measuring trends in contraceptive use. In: Current and Future Fertility Research - The Proceedings of a Seminar Sponsored by the Canadian Committee for Fertility Research. Family Planning Division, Health and Welfare Canada, 1982:2-15.

128. Hosmer DW, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. Communications in Statistics A 1980;9:1043-69.

129. Vermesh M. Conservative management of ectopic gestation. Fertil Steril 1989;51:559-67.

130. Garcia AJ, Aubert JM, Sama J, Josimovich JB. Expectant management of presumed ectopic pregnancies. Fertil Steril 1987;48:395-400.

131. Wartenberg D, Agamennone VJ, Ozonoff D, Berry RJ. A microcomputer-based vital records data base with interactive graphic assessment for states and localities. Am J Pub Health 1989;79:1531-6.

Appendices

Appendix A

Study conditions set by Saskatchewan

The following conditions were set by the Saskatchewan Hospital Services Plan prior to release of the data:

1) [The researcher] will be responsible for all costs incurred in producing data for the study.
2) Patient confidentiality will be respected. Only aggregate results will be reported in any publication.
3) Only data on patients separated from hospital with a diagnosis of ectopic pregnancy and an appropriate control group will be released. This data will not include individual identifiers.
4) The data will be provided solely for the study of risk factors for ectopic pregnancy. Any further use of the data will require separate approval from SHSP.
5) Prior to any public release of the study results, a written summation of the work will be presented to SHSP to determine if confidentiality will be threatened by publication.
6) Any written or verbal release of the study results should indicate that the responsibility of the study and its conclusions are those of the author and not SHSP.

dated January 26, 1987

Appendix B

Structure of SHSP database and translated files

Table 24. Format of records received from SHSP

Variable Name	Start Column	Width	Description
ID	1	8	Unique anonymized identifier (sequential)
YOB	9	2	Year of birth
MOB	11	2	Month of birth
AGRP	13	2	Age group code
ICD1	15	4	Primary ICD code without decimal
CLIST1	19	4	Primary CLIST code
ICD2	23	4	Secondary ICD code
CLIST2	27	4	Secondary CLIST code
SEPY	31	2	Year of separation
SEPM	33	2	Month of separation
SEPD	35	2	Day of separation
OPCODE	37	4	CCP procedure code
OPDAY	41	1	Days from admission to first procedure
DAYS	42	4	Days in hospital
ADMTYPE	46	1	Admission type 1=emergency 2=urgent 3=elective
SEPTYPE	47	1	Separation type 1=discharged 2=discharged without authority 3,4=death 5=transfer to hospital 9=other
CASE	48	1	Case status 1=Case 2=Control
MATCH	49	8	Case ID for which control selected
I89	57	1	ICD revision for codes (8 or 9)

All fields are numeric except for ICD1, ICD2, and OPCODE, which may have a single letter prefix

Table 25. Translated file format

Variable Name	Start Column	Width	Description
ID	1	4	Unique ID
CASE	5	1	Case=1, Control=2
The following fields relate to index separation			
AGRP	6	2	Age group
ICD1	8	4	ICD-9 code for primary diagnosis
ICD2	12	4	ICD-9 code for secondary diagnosis
OPCODE	16	4	Operation code
D1	20	2	my index for primary diagnosis
D2	22	2	my index for secondary diagnosis
OP	24	2	my index for operation code
SEPY	26	2	Year of separation
SEPM	28	2	Month of separation
SEPD	30	2	Day of separation
ADMTYPE	32	1	Type of admission
SEPTYPE	33	1	Type of separation
OPDAY	34	2	Days to operation
AGE	36	2	Age in whole years
DAYS	38	3	Days in hospital
The following fields relate to separations prior to index separation			
AGEFIRST	41	2	Age at first recorded separation
NSEP	43	2	Total number of separations
G	45	2	Gravidity (includes index)
P	47	2	Parity (excludes index)
G1	49	2	Age at first known pregnancy
GN	51	2	Age at last known pregnancy
The following fields are repeated for each historical diagnosis of interest (see list in Appendix C)			
	53 ...	6	
COND(i)		2	Number of separations
AGEF(i)		2	Age at first known separation
AGEL(i)		2	Age at last known separation

Appendix C

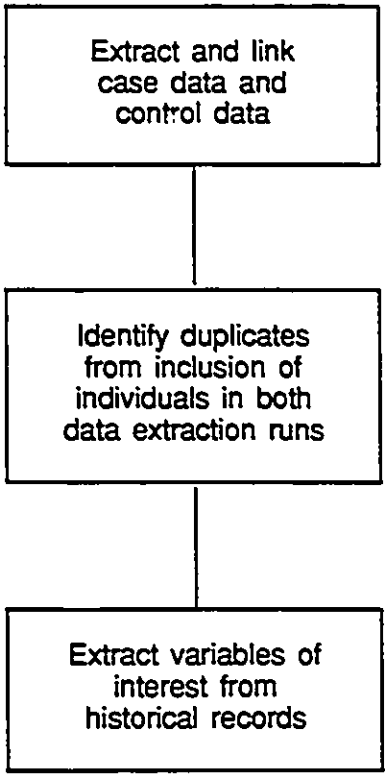

Figure 10. Overview of data transformation for case-control study

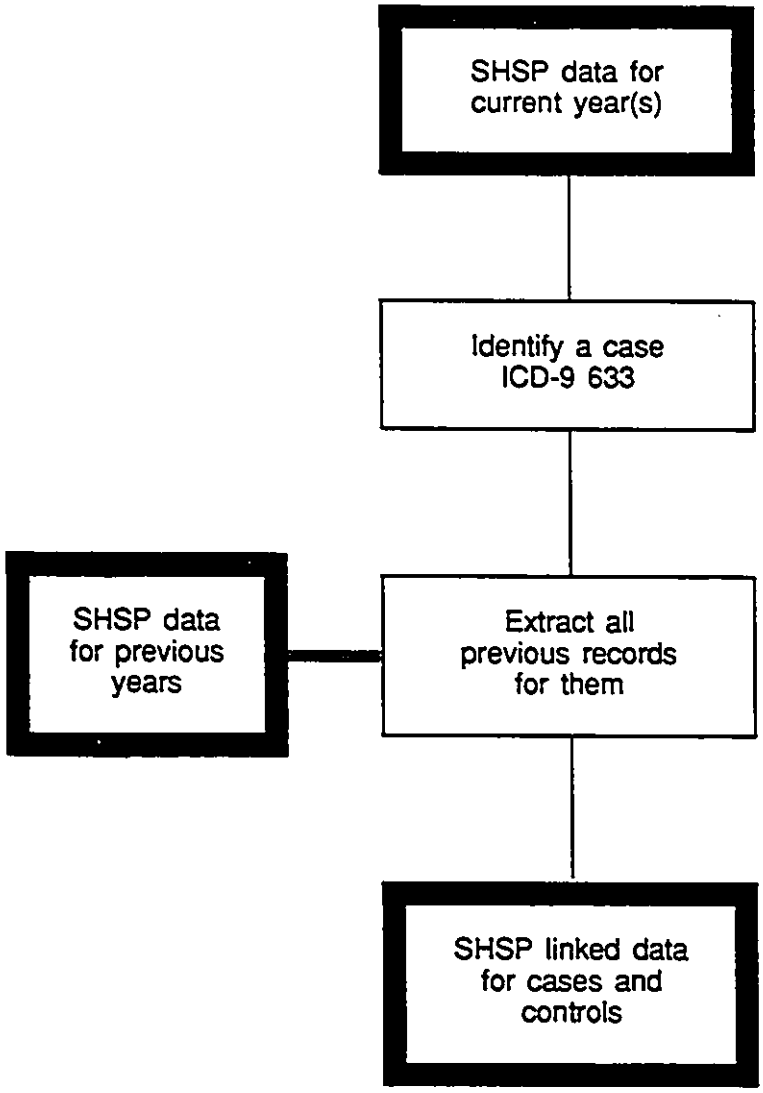

Figure 11. Flow chart of SHSP extraction of case and control data

——— Flow of data ▬▬▬ Record linkage

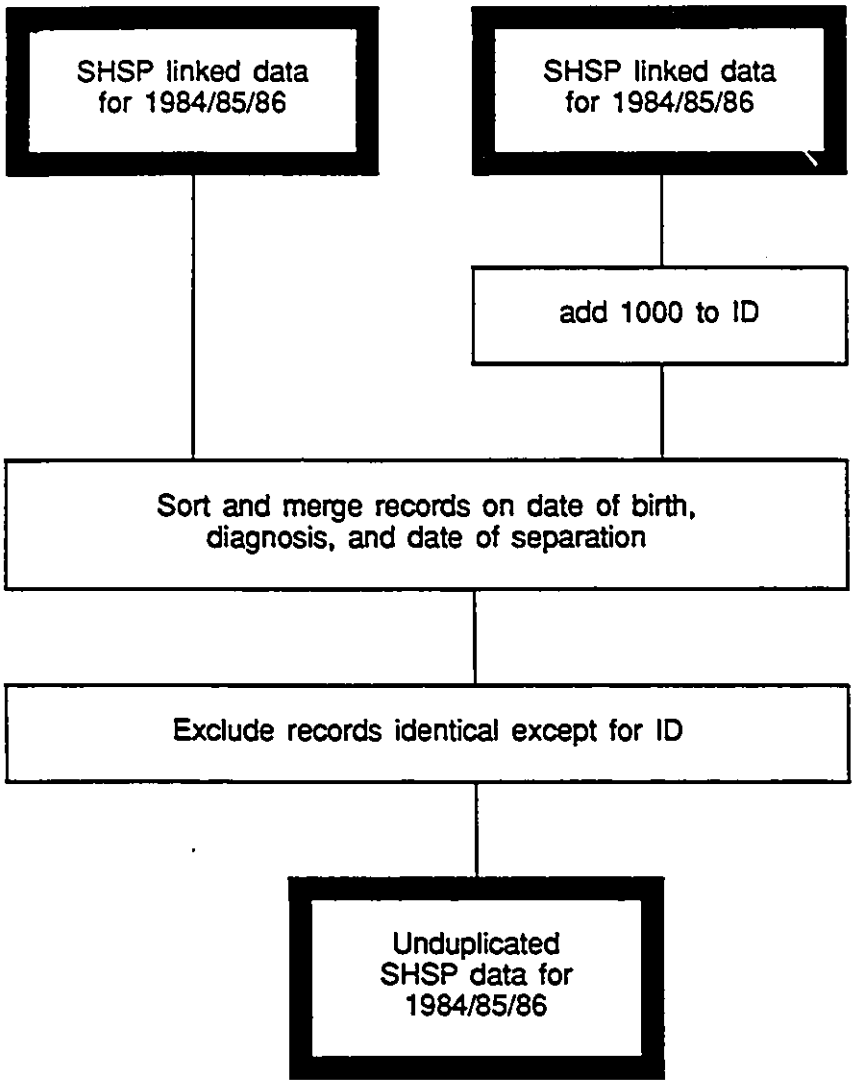

Figure 12. Flow chart for exclusion of duplicate cases or controls

An individual may be included as a case from one source data set and as a control in the other, since records will differ on the assigned case or control classification.

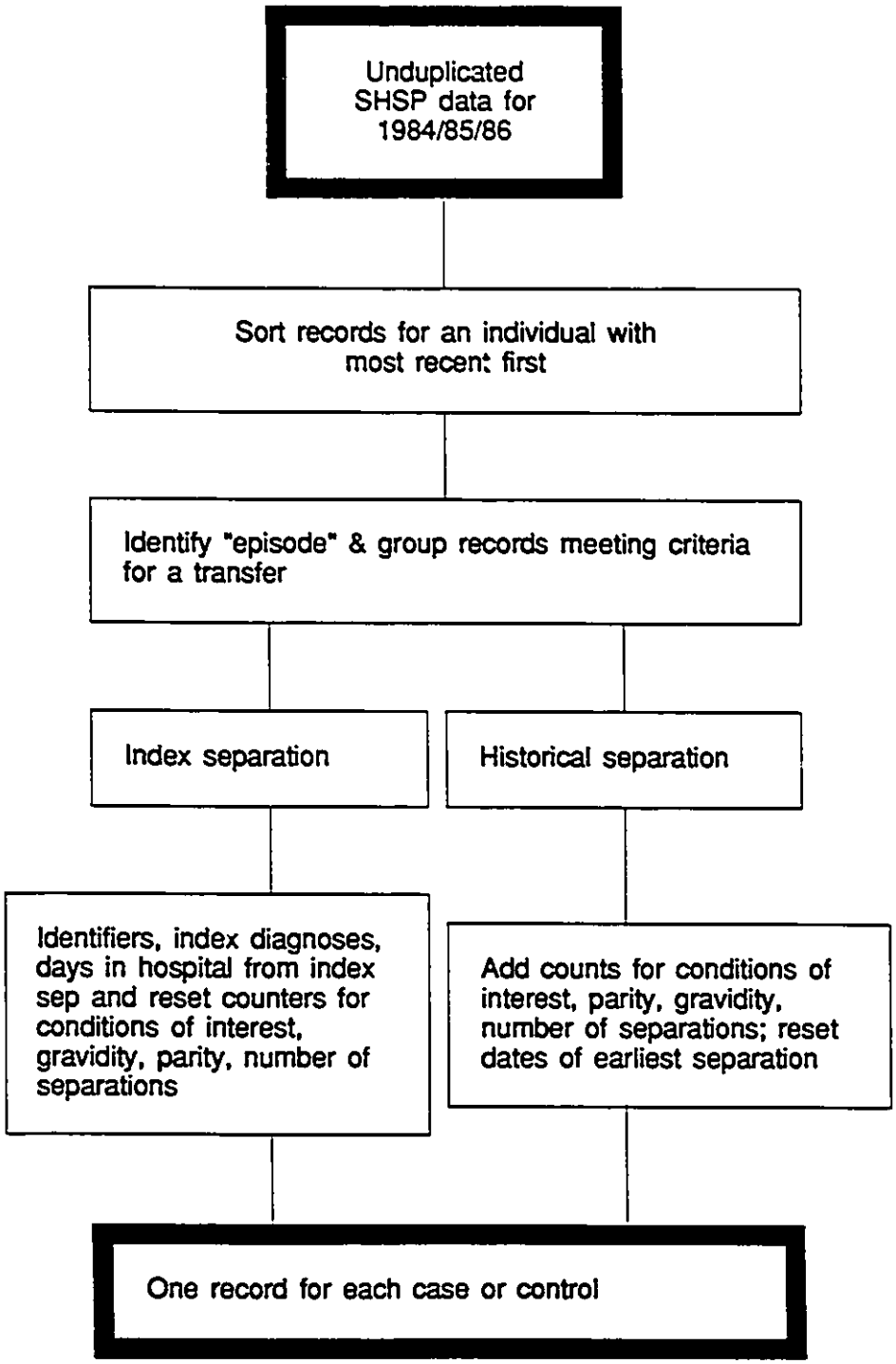

Figure 13. Data transformation algorithm

Table 26. ICDA-8 diagnosis codes for each condition of interest

Condition	ICDA-8 codes
Acute PID	612, 614, 616.0
Chronic PID	613, 616.1
STD,infection	091-099, 620.9, 622.0
Endometriosis	625.3
Abortion, induced*	640-642
Infertility	628.0
IUD in situ	-
Sterilization	-
Tubal repair	-
Ectopic pregnancy*	613
Abortion, spontaneous*	643-645
Tubal surgery	-
Caesarean section††	-
Delivery††	650-662

*contributes 1 to gravidity
†contributes 1 to parity

Table 27. ICD-8 Procedures & Operations codes

Condition	ICDA-8 codes
Acute PID	-
Chronic PID	-
STD,infection	-
Endometriosis	-
Abortion, induced	746-749
Infertility	-
IUD in situ	-
Sterilization	682
Tubal repair	684
Ectopic pregnancy	-
Abortion, spontaneous	-
Tubal surgery	681-685, 689.0
Caesarean section	770-779
Delivery	750-769, 780-789

Table 28. ICD-9 diagnosis codes for each condition of interest

Condition	ICD-9 codes
Acute PID	614.0, 614.2, 614.3, 614.8, 614.9
Chronic PID	614.1, 614.4, 614.6, 614.7
STD,infection	091-099, 615.0, 615.9, 616.0
Endometriosis	617
Abortion, induced	635, 636, 638
Infertility	628.2, 628.8, 628.9
IUD in situ	V45.5
Sterilization	V25.2
Tubal repair	V26.0
Ectopic pregnancy	633
Abortion, spontaneous	630, 631.0, 632, 634, 637
Tubal surgery	-
Caesarean section	-
Delivery	650-653, 660-669

Table 29. CCP Procedures & Operations codes

Condition	ICD-9 codes
Acute PID	-
Chronic PID	-
STD,infection	-
Endometriosis	-
Abortion, induced	87.0-87.2
Infertility	-
IUD in situ	-
Sterilization	78.3-78.4, 78.53
Tubal repair	78.6
Ectopic pregnancy	-
Abortion, spontaneous	-
Tubal surgery	78.0-78.5, 78.8, 78.9
Caesarean section	86.0-86.2, 86.8-86.9
Delivery	84.0-84.9, 85.6-85.9, 87.6-87.9

Appendix D

Birth order by maternal age for Canada

Table 30. Hundreds of live births by birth order and maternal age - Canada 1986

	Birth order								
Age group	1	2	3	4	5	6	7	8	Total
15-19	179	31	3.8	-	-	-	-	-	214
20-24	542	294	75	14	2.7	-	-	-	928
25-29	586	567	211	53	12	4	1.2	-	1434
30-34	221	326	181	59	16	6.4	2.6	1.1	813
35-39	49	75	56	25	9.5	4.2	2	1.2	222
40-44	5	6.2	5.6	3.7	2	1.1	-	-	23
Total	1582	1299.2	532.4	154.7	42.2	15.7	5.8	2.3	3634

- indicates less than 50
Source: Statistics Canada. Vital Statistics: Births, 1986