CHAPTER 1

Introduction

1.1 Background

The incidence of failure of a method for contraception is generally a matter of great interest to any person who uses, or whose sexual partner uses, that method. Not surprisingly, there is a large volume of research into the failure rates of all of the temporary methods for human contraception. However, there is a much more modest literature on the cumulative incidence of failure and annual failure rates of permanent sterilisation, particularly failures of female tubal sterilisation - often called "tubal ligation", but including any means for occluding or interrupting the Fallopian tubes by surgical means. This is somewhat surprising given that female tubal sterilisation remains one of the most popular and widely used means of contraception.

The Australian Study of Health and Relationships, which was conducted between May 2001 and June 2002 using a representative sample of 9,134 women aged 16 to 59 years, found that of the two-thirds of respondents who reported using some form of contraception, 22.5 per cent relied on tubal ligation or hysterectomy (these were not further distinguished), which was second in popularity only to oral contraceptives.[1] The proportion of women in the 40-49 year age group who relied on tubal ligation or hysterectomy was 33 per cent, which was second in popularity only to vasectomy in the partner (34.6 per cent).

In the 1995 United States Survey of Family Growth, conducted by the US National Center for Health Statistics, surgical sterilisation was the method of choice in 29.4, 40.9 and 49.8 per cent of women who were using any form of contraception in the 30-34, 35-39 and 40-44 year age groups respectively.[2] Female tubal sterilisation also remains an extremely popular method of contraception in developing and transitional countries, although the high prevalence of HIV infection in many parts of Africa and

some parts of South and South-East Asia has resulted in the promotion of barrier methods of contraception over surgical sterilisation, even for older women who have completed their families.[3] Data on the number of tubal sterilisations performed in New South Wales will be presented later.

Extant investigations of the incidence of tubal sterilisation failure fall into four broad groups: a) smaller cohorts based on personal or hospital case series, with variable degrees of follow-up; b) large observational cohorts formed by pooling subjects drawn from many hospitals, with long-term active follow-up; c) randomised controlled trials (RCTs) designed to assess the safety and effectiveness of different sterilisation procedures; and d) case-control studies of patients presenting with ectopic pregnancy. Case series typically lack internal validity due to the use of informal and inconsistent follow-up mechanisms, and almost always have poor generalisability due to unquantified and uncontrolled selection biases. Pooled cohort studies with active follow-up tend to be very expensive and thus have not been replicated widely, nor repeated as surgical practice and methods change. Similarly RCTs tend to be expensive and are now performed only for newly introduced methods. Depending on the recruitment frame and inclusion or exclusion criteria used, they may also have limited generalisability. That leaves case-control studies of ectopic pregnancies as the only (relatively) cheap and easily replicated study design reported in the literature to date, albeit one which is restricted to a less common mode of failure and which therefore cannot provide an estimate of the overall incidence of failure.

Thus the primary motivation for the study described in this thesis was to assess whether a) a very inexpensive study design, using routinely-collected data and record

linkage methods, could be used to derive lower-bound estimates for cumulative incidence of failure for female surgical sterilisation in a large, geographically-defined population and; b) whether those estimates are comparable with estimates of cumulative incidence of failure reported in the literature.

1.2 Scope

This thesis reviews extant studies of failure incidence of surgical tubal sterilisation, and then estimate a lower bound for the cumulative incidence of failure of these procedures using probabilistically linked records of admissions to New South Wales hospitals. In this study it was only possible to estimate a lower bound on the cumulative incidence of failure due to the necessarily incomplete nature of the follow-up process using record linkage. A secondary aim was to explore the effect of various factors on the estimated failure incidence, using survival modelling techniques. However, it is not possible to undertake a definitive analysis of the effect of such factors because of unquantifiable biases which are likely to exist in the linked data source. These putative biases will be discussed in more detail later in Chapter 3.

1.3 Organisation of this thesis

Chapter 1 comprises this introduction.

Chapter 2 is a literature review. A comprehensive review of the extensive biomedical and demographic literature on the use and incidence of failure of all methods of contraception is beyond the scope of this thesis. Rather, this literature review will focus on the following areas within the wider domain of contraception research:

- Description of the basic reproductive anatomy, physiology and surgical techniques relevant to tubal sterilisation.

- Reviews of methodological issues in studies of contraception failure
- Scientific papers which report incidence of failure of tubal sterilisation, including studies which focus on the incidence of ectopic pregnancy following tubal sterilisation.
- Scientific papers which examine other sequelae of tubal sterilisation, including menstrual disturbances, increased likelihood of hysterectomy, and death.

Chapter 3 describes the methods used to probabilistically link hospital admission records in order to determine whether tubal sterilisation procedures may have failed. A brief review of record linkage techniques and methods as described in the literature appears is also included in this chapter, as is a discussion of the weaknesses of and possible biases present in the linked data.

Chapter 4 presents and considers the main results of this study.

Chapter 5 presents a proportional-hazards analysis which explores one hypothesis, that the incidence of sterilisation failure is associated with the volume of sterilisations performed in each hospital.

Chapter 6 provides a summary of the study and its findings in the context of other studies of tubal sterilisation failure, briefly examines the potential social, psychological and health service implications of those results, and provides some recommendations for future studies and mechanisms for monitoring incidence of surgical sterilisation failure.

Computer programme code, intermediate results and other technical details appear in a series of appendices. Finally, two published, peer-reviewed papers related to the use of record linkage for disease surveillance and health service evaluation, written or co-written by the author during the course of candidacy for the degree for which this thesis is submitted, appear as addenda.

CHAPTER 2

Literature Review

2.1 Goals

The primary goal of this literature review is to identify and appraise studies that have attempted to estimate the incidence of failure of tubal sterilisation procedures, in order to provide comparison points and a quantitative context for the cumulative incidence of failure estimated by the current study. The review is primarily descriptive in nature, and no quantitative meta-analysis is attempted, primarily because the epidemiological quality of the majority of the published studies was insufficient to make such an analysis worthwhile. However, relevant retrieved studies were assessed using the critical appraisal frameworks for RCT, case-control and cohort studies published by UK National Health Service Public Health Resource Unit.[4]

2.2 Methods

2.2.1 Search

Suitable English language publications were initially identified through a series of searches of the PubMed bibliographic database, operated by the US National Center for Biotechnology Information.[5] PubMed incorporates the MEDLINE bibliographic database, operated by the US National Library of Medicine, as well as many journals not indexed by MEDLINE. PubMed queries were formulated using the strategies recommended by Ebbert *et al*.[6]

A primary search was undertaken using the MeSH (Medical Subject Headings) "Sterilization, Tubal" and "Sterilization, Sexual" (which was the indexing term used prior to 1972). In addition, text word searches were performed for "contracept* and failure*" – that is, citations containing words starting with "contracept" or "failure". Finally a search for methodological papers was performed using the MeSH heading

"Contraception" qualified by the MeSH subheading "/methods". Papers were assessed for possible relevance based on their titles and abstracts where available, using the criteria listed in section 2.2.2 below.

PubMed and MEDLINE do not, in general, index medical monographs. Therefore searches of the library catalogues of the Universities of Sydney and NSW were made, as well as the Amazon online bookseller (http://www.amazon.com) and the Google index of World Wide Web sites (http://www.google.com).

Finally, the reference lists of all publications identified through bibliographic searches and subsequently retrieved were scanned for further relevant papers, and some of these references were also retrieved in an iterative fashion. Recent work by Kuper *et al*. found that, for a review of literature on cardiovascular disease and depression, citation tracking tended to find additional relevant papers in higher impact-factor journals that were missed by bibliographic database searches (that is, PubMed/MEDLINE) alone.[7]

Several of the historical literature reviews which were located cite hundreds of case series reports stretching back to the early 20th century, and no attempt was made to retrieve these citations, although copies of all relevant studies reporting cumulative incidence of failure or failure rates, published since 1980, were sought. A small number of potentially relevant papers were not able to be obtained at reasonable cost – in these cases, only the abstract was reviewed, and this fact noted.

2.2.2 Relevance criteria

Apart from documents providing background information on anatomy, physiology, surgical techniques, epidemiological methodology and long-term non-failure sequelae of tubal sterilisation, the sole criteria for inclusion in the review was reporting of a quantified absolute incidence or incidence rate or failure, or a rate or odds ratio, for tubal sterilisation procedures, where failure is defined as any form of conception, including ectopic pregnancies.

2.3 Biomedical foundations

2.3.1 Review of anatomy and physiology of the Fallopian tubes

Knowledge of the basic anatomy of the female reproductive organs is assumed. The following account of anatomy and ovum transport is taken from Moawad and Hafez and from Hacker and Moore.[8,9]

The Fallopian tubes (oviducts) are bilateral muscular tubes about 10 cm in length which connect the uterine cavity with the peritoneal cavity. They have four functional segments: the *intramural* or *interstitial part*; the *isthmus*, which is a narrow proximal portion of the Fallopian tube; the more dilated and distal *ampulla*, where fertilisation of the ovum usually takes place, and the funnel-shaped opening to the abdominal cavity, the *infundibulum*, which has fringe-like *fimbriae* which lie above and laterally to the ovary. The oviducts are lined with a mucosa comprising a mixture of secretory, ciliated and intercalary "peg" cells. The mucosa itself is arranged in a series of high, branched folds which almost entirely fill the lumen of the oviduct – egg transport occurs between these folds, rather than down the centre of a hollow tube. Ciliated cells are most common near the fimbriae, and decrease in number towards the intramural part of the oviduct. Following ovulation, eggs

surrounded by cumulus cells are transported from the ovarian surface to the infundibulum by a combination of beating of the ciliated cells, contractions of the fimbriae, together with a sucking action caused by contractions of the muscular layer of the Fallopian tube (the *myosalpinx*). An egg reaches the ampullary-isthmic junction about 30 hours after ovulation, after which transport through the isthmic portion is quite rapid – the total time which the ova spend in the oviducts appears to range between 45 and 80 hours.

The oviducts also serve to transport sperm to the ampullary region, where fertilisation usually occurs. The primary method of sperm transport appears to be a distal secretory flow of oviductal fluid, which also serves to capacitate (activate) sperm. Sperm self-motility only appears to be important once the sperm have reached the ampullary part of the oviduct.

The relative size of the ovum and sperm are also important when considering the mechanisms and modes of failure of female surgical sterilisation. A mature ovum is between 120 and 150 microns in diameter, whereas a normal human spermatozoon has a head width of 2.5 to 3.5 microns and a tail length of 5 to 7 microns.

2.3.2 Methods for female surgical sterilisation

Tubal sterilisation was first described in 1881 by Lungren as an incidental procedure to caesarean section.[10] Since then, over 100 different surgical sterilisation techniques involving the Fallopian tube have been described. Only the more popular procedures will be described in this brief review, which has been synthesised from chapters in three texts covering operative gynaecology.[11,12,13]

Techniques can be broadly classified into laparoscopic methods, including electrocoagulation (unipolar or bipolar), Silastic band (Yoon) or spring-loaded clip (Hulka-Clemens, Rocket or Filshie), and open techniques (often via a mini-laparotomy) such as the Pomeroy, Irving, Uchida and Kroener methods.

Unipolar electrocoagulation involves grasping the mid portion of the tube with forceps and applying current for approximately five seconds until a section of the tube can be avulsed. The proximal stump of the tube is then further cauterised. Bipolar electrocoagulation is similar, except that the two jaws of the grasping forceps form the cathode and anode, which helps reduce the likelihood of inadvertent thermal injury to surrounding structures (particularly bowel).

The Yoon band technique (also known as the Falope-ring) relies on a small ring made of silicone elastomer (Silastic). The Fallopian tube is grasped through a laparoscope at the ampullo-isthmic junction and the band is pushed over a knuckle of tube, which then undergoes ischaemic necrosis over the next few days. The Hulka-Clemens clip consists of two small serrated Silastic jaws held together by a metal clip. It is also applied via a laparoscope in a similar manner to the Yoon band. The Hulka-Clemens and closely related Rocket clips have now largely been replaced by the Filshie clip, which differs in having a fully-articulated hinge and a metal locking mechanism to keep the jaws of the clip together after application.

There are several open techniques which were popular in the past, and which gave rise to the widely-used term “tubal ligation”. The Irving method involves double ligation of the tube, and resection of its middle portion. The proximal stump is then embedded on the anterior surface of the uterus in a pocket formed in the

myometrium. The Uchida method involves ligation and division of the tube within the mesosalpinx which covers it. The proximal end of the distal part of the now divided tube is exteriorised, so that it is open to the abdominal cavity, and the mesosalpinx is repaired, leaving the distal end of the proximal part of the divided tube inside – thus the mesosalpinx is used as an additional barrier between the two divided ends of the oviduct.

The Pomeroy method was also very widely used. It is similar to the Irving procedure except that a knuckle of tube is formed using a single ligature, and the knuckle excised. The intact mesosalpinx then tends to keep the severed ends of the tubes widely separated after healing. The Kroener method involves complete removal of the outer third of the oviduct and fimbriae – the stump of the tube is secured with two separate ligatures.

Numerous methods for blocking the uterotubal ostia through the use of a hysteroscope have been described, including chemical fulguration and plugging the tubes with silicone or pre-formed mechanical plugs. None of these techniques have become popular due to the relatively complicated and expensive equipment required, technical difficulties in performing the procedures, and a high failure incidence and very high ectopic pregnancy rate.[14,15,16,17] These methods will not be considered specifically in this work.

For the sake of completeness, it must be mentioned that hysterectomy (with or without bilateral oophorectomy) also results in sterility. However, it seems unlikely that hysterectomy and/or bilateral oophorectomy are now used very much where contraception is the sole indication – at least not in the last two decades. No studies

of female sterilisation done since the early 1980s have included hysterectomy in their scope. A few earlier studies do report hysterectomy as a distinct means of surgical sterilisation, albeit one considerably less popular than tubal sterilisation.

2.4 Methodological issues in the estimation of the incidence of contraception failure

The following discussion is based primarily on review articles by Trussel *et al.*, by Farley and by Trussel.[18,19,20]

2.4.1 Types and causes of failure

Some studies of contraception efficacy attempt to distinguish between *method* failure – due to inadequacy of the method itself - and *user* failure – due to user error in applying (or failing to apply) the method. In the context of female surgical sterilisation, there is no scope for user error when the user is defined as the woman or couple involved (note that this is not true of male surgical sterilisation, because abstinence or other forms of contraception must be used for some time after the surgical procedure).

However, it can be argued that method failures of female surgical sterilisation can still be partitioned into failures due to an intrinsic inadequacy of the particular surgical method used, and failures due to incorrect application of that method. The latter may be due to factors such as poor surgical technique (such as mistaken ligation of the round ligament instead of the oviduct), equipment failure (for example, poor calibration of a clip applicator, resulting in incorrect tensioning of the applied clip on the oviduct), or a failure to exclude either uterine pregnancy or luteal

phase pregnancy prior to surgery.[21] Many investigators have speculated on the causes of method failure, but few have been able to gather objective evidence.

Lipscomb, Spellman and Ling examined the effect of day-of-surgery pregnancy testing on incidence of failure of tubal sterilisations due to luteal pregnancies.[22] In 401 laparoscopic tubal ligation procedures, pregnancy was excluded by urinary hCG (human chorionic gonadotrophin) testing at the pre-operative examination conducted some time in the two weeks prior to surgery. Seven subsequent pregnancies occurred in which the timing of conception, based on ultrasound appearances, was apparently prior to the surgery. The procedure for excluding pregnancy was changed to a more sensitive ELISA (enzyme-linked immunosorbent assay) blood test on the day of surgery. In 605 subsequent procedures, eight operations were cancelled due to a positive hCG test, and no luteal phase failures were subsequently detected. This study suggests that undetected luteal phase pregnancies may be a significant cause of apparent failures of tubal sterilisation, at least in settings in which day-of-surgery pregnancy testing is not practiced. It is debatable whether the failure to detect such pregnancies prior to surgery should be regarded as a method failure. Regardless, for a retrospective, population-based study such as the current one, the exclusion of possible luteal phase pregnancies is problematic. If data on the estimated gestational age were routinely available for apparent failures resulting in a birth or termination, then those pregnancies in which conception appeared to pre-date the sterilisation procedure could be censored. However, such information is rarely available on a population basis, especially for ectopic pregnancies, and in this study, not for any pregnancies. Options for dealing with this issue are discussed further in Chapter 3.

2.4.2 Failure modes

Apart from undetected luteal phase (and possibly uterine) pregnancies, any form of conception following a tubal sterilisation procedure unambiguously constitutes a failure. This was implicit in all of the studies examined in the course of this literature review. A few studies explicitly listed the events which they considered to denote a failure, specifically: delivery of a neonate (alive or stillborn), ectopic pregnancies, and spontaneous or induced abortions. None of the studies reviewed specifically mentioned gestational trophoblastic disease (that is, complete and partial hydatidiform moles, choriocarcinomas, and invasive moles), perhaps due to the rarity of these conditions.

2.4.3 Modes of withdrawal and censoring events

Studies of most types of contraception are designed to estimate not just the incidence and incidence rates of (unwanted) pregnancy, but also the acceptability of the contraceptive method itself – hence, several reasons for withdrawal from a study are usually considered. For example, in a study of intrauterine devices (IUDs), withdrawals (failures) may be caused by pregnancy, expulsion of the device, infection or request for removal due to excessive bleeding, pain or other reasons. However, in all studies of female surgical sterilisation reviewed as part of this thesis, only conception indicated by uterine or ectopic pregnancy or abortion was regarded as a failure. Surgical reversal of sterilisation was either not mentioned, or was considered a censoring event, rather than as a separate study end-point. In other words, all extant studies have focused on (presumably unwanted) conceptions as the outcome of interest, rather than on requests for reversal of sterilisation. The same approach was taken in the current study, with an indication of attempted reversal treated as a censoring event. There is a small body of studies, not reviewed here, on

both the technical outcome of reversal procedures and on the social and psychological reasons leading women or couples to seek reversal. Such studies are outside the scope of this thesis.

A fundamental requirement in all survival (time-to-failure) analyses is that censoring events are *non-informative* – that is, a censored observation is one whose value is incomplete due to random factors (with respect to the outcome of interest) for each subject.[23] It could be argued that some types of sterilisation procedure have a higher risk of causing side effects, such as dysfunctional uterine bleeding or other menstrual disturbances, which in turn leads to a higher risk of the subject undergoing a censoring event, such as a hysterectomy. It is quite likely that the type of sterilisation procedure will also have some influence on the time-to-failure. Thus, it is possible that an association may exist between censoring due to hysterectomy and the time-to-failure: such censoring could be *informative*. If this were the case, hysterectomy events should be treated as *right truncation* of that case, not right censoring. However, truncated data are generally not accommodated by any of the usual survival analysis methods, and such observations need to be excluded, which wastes data and may introduce additional biases. In the current study, hysterectomy was treated as a censoring event rather than a truncating event, and the theoretical possibility that it may be informative was acknowledged but not further explored.

2.4.4 Summary estimates of incidence of contraception failure

2.4.4.1 The Pearl rate

The Pearl rate, first proposed in 1933, was used to summarise incidence of pregnancy in most contraception studies up until the late 1960s - it is simply the number of events divided by the number of months of exposure times 1200, expressed as an

incidence rate per 100 woman-years of exposure.[24] It is an adequate summary if the underlying risk of an event is constant throughout the follow-up period. Potter pointed out that this is rarely the case: post-partum, the risk of pregnancy is low due to lactational suppression of ovulation, and that thereafter the risk varies with frequency and timing of intercourse.[25] Furthermore, those subjects with a higher risk of pregnancy (due to method failure, higher intrinsic fertility or other reasons) will tend to become pregnant earlier in the follow-up period, while the remaining subjects necessarily have a lower risk of failure. However, this defect is unlikely to be of importance in studies in which failures are relatively rare, as is generally the case with female surgical sterilisation failures – the small number of failures do not materially affect the risk in the remaining, much larger, population-at-risk. Nor is the Pearl rate a measure of cumulative incidence, which would require multiplying it by the length of time of follow-up. However, follow-up time tends to vary by subject, and thus the majority of studies of female surgical sterilisation failure since the 1970s have used life table or other survival analysis techniques.

2.4.4.2 Life-table techniques

As noted by Farley, the main advantage of life-table techniques is the ability to right-censor data: that is, the ability to handle subjects who have been lost to follow-up, or who have experienced an event which signifies end-of-risk before the end of the study follow-up period.[18] In the context of contraception studies, end-of-risk events usually include completion of menopause, or hysterectomy and/or bilateral oophorectomy. Strictly speaking, cessation of heterosexual activity (and/or attempts at assisted conception) should also be considered as a censoring event, because *in vivo* conception is impossible without the presence of spermatozoa in the female

reproductive tract. However, relatively few contraception studies consider these as a censoring event due to the difficulty in collecting reliable data.

The basic formulation used for "observed" (or "cohort") life-table analysis was first given by Berkson and Gage in 1950 and was generalised to handle multiple causes of failure by Potter in 1967.[25,26] The notation used below is taken from Armitage and Berry.[27]

The observed life-table is constructed with one row for each time interval, from time x to time x+1, relative to the intervention or treatment of interest – that is, subjects may enter the study on different calendar dates, but "study time" is measured relative to their entry into the study. It should be noted that traditional cohort life-table analysis and related analysis techniques (such as the Kaplan-Meier estimator) do not permit "left censoring" – that is, all subjects are assumed to be under observation from the time of the intervention or treatment of interest until failure occurs or until they are right censored due to loss to follow-up or study completion. There was no left censoring in the current study.

The estimated probability of failure, q_x, in the interval x to x+1 is number of subjects in whom a failure is observed, d_x, divided by the adjusted number at risk, n'_x, given by $n'_x = n_x - w_x / 2$ where n_x is the number still at risk and still under observation at the beginning of the interval, and w_x is the number who are withdrawn (right censored) in that interval. The rationale for the adjusted number at risk is that withdrawals are at risk for only part of the interval. Although it is desirable to account for the exact proportion of the interval for which they were at risk, in practice it is usually adequate to assume that each withdrawal was effectively at risk

for half of the interval.[27] Of course, this assumes that the risk of failure is uniform throughout the interval. If the intervals are sufficiently small, this assumption is justified, but when intervals of one year or more are used, it may not be the case. Indeed, when constructing life tables, demographers routinely use an adjustment for the first year of life to account for the greater death rate immediately after birth. There is evidence from a large US pooled cohort study (reviewed in more detail below) that risk of failure declines over the course of several years following female surgical sterilisation.[28] No empirical information regarding variation of risk in the first few weeks or months following the sterilisation procedure was found in the course of this literature review.

The cumulative probability of failure from entry into the study to interval x is:

$$\Phi_x = 1 - \prod_{u=1}^{x} (1 - q_u)$$

Greenwood[29] described the calculation of estimated variance for each interval failure rate, based on binomial sampling:

$$\text{var}(\Phi_x) = \Phi_x^2 \sum_{u=1}^{x} \frac{d_u}{n'_u (n'_u - d_u)}$$

Due to the use of an asymptotic normal approximation to the binomial distribution in this formula, it is possible for confidence limits calculated using this estimate of variance to lie outside the range 0 to 1. To avoid this, Kalbfleish and Prentice[30] suggested the use of a double logarithmic transformation, so that 95 percent confidence limits for the cumulative probability of failure become:

$$\Phi_x^{\exp(\pm 1.96\sqrt{\text{var}(\Phi_x)}/(-\Phi_x \ln(\Phi_x)))}$$

Armitage and Berry point out that there are two further assumptions implicit in these calculations: a) although a single value of q_x is calculated for each interval, all the subjects contributing to that interval failure rate experienced their failure at different

absolute times (although at approximately the same time relatively to their entry into the study) – thus interval failure rates must be assumed to remain constant over calendar time, and b) that subjects who are censored have the same risk of failure while they were still considered to be under observation as those who remain in the study. As will be discussed further in the next section, the current study relied on a form of passive follow-up through probabilistic record linkage. As a result, although many subjects were lost to follow-up, there is no way of knowing who they are, and thus they cannot be censored at the appropriate time. Trussell provides an extensive discussion of this issue, and argues that, in the context of typical studies of contraception effectiveness, if one assumes women are lost to follow-up (LFU) because they became pregnant and thus dissatisfied with the study organisers, then only an upper-bound estimate of the overall failure rate (or rather, the cumulative incidence of failure) can be calculated, whereas if it assumed that none of the women who are LFU became pregnant, then only a lower-bound estimate is possible.[18] In this study, those subjects who were silently lost to follow-up continue to contribute to the denominator of the interval failure rates, but not to the numerator, regardless of whether had a higher or lower risk of failure than those who effectively remain under (passive) follow-up. It is for this reason that only a lower bound estimate could be calculated.

One other important analytical issue canvassed by both Farley and Trussell relates to the use of multiple-decrement life-tables in studies with multiple causes of withdrawal (failure): that the reporting of "net" pregnancy failure rates from multiple-decrement life tables may be biased if there are significant numbers of withdrawals due to other reasons. This point can be appreciated if one imagines two populations of women who experience identical conditional probabilities of failure

throughout a study of some contraception method. If half the women in the second population discontinue use of the contraception method, and are thus censored, while all of the women in the first group continue to use it, it is clear that the overall number and thus the proportion of failures will be higher in the first group.

However, as already discussed, this issue is not relevant to studies of surgical sterilisation, because "discontinuations" of the method by attempted reversal are rare and are therefore not explicitly accounted for in the studies reviewed. Thus any kind of pregnancy is almost universally considered to be the only outcome of interest in any failure analysis.

Actuarial life tables are now rarely used time-to-event analysis in epidemiological studies – the Kaplan-Meier estimator is now much more commonly used.[31] The Kaplan-Meier estimator is also known as the product-limit estimator because it is the limiting case of the survival estimate obtained from actuarial life tables, with the time periods shrunk so that an entry is effectively made for each distinct failure time. The formula for survival at time *t* for the Kaplan-Meier estimator is essentially identical to the formula for the actuarial life table, and Greenwood's formula with the Kalbfleish and Prentice transformation and can also be used to calculate variance and hence confidence limits.

Only two of the studies of female surgical sterilisation reviewed here, both undertaken in the 1990s, made any use of continuous time survival analysis methods (such as product-limit estimators, or parametric or proportional hazards regression modelling), in order to compare the performance of different methods. This probably

reflects the fact that researchers did not have ready access to, or were not familiar with, software to undertake such analyses until the last decade.

2.5 Studies which report failure of female surgical sterilisation.

Studies are reviewed here in approximately chronological order. The review does not attempt to be exhaustive for relevant papers published prior to about 1990, partly due to the sheer number of such papers, and partly because of the serious methodological problems which most of them possess - primarily ill-defined follow-up mechanisms and periods, and consequently under-estimated numerators and over-estimated denominators (due to lack of right censoring of subjects no longer at risk and/or no longer under effective follow-up). Of course, the same criticisms can be levelled at the current study, which is why only the estimation of a lower bound is claimed. For the same reasons, estimates of the incidence of failure reported in earlier papers should generally be regarded as lower bounds for the actual failure incidence. Greater attention is paid in the following review to larger, better designed studies of sterilisation failure, of which there are only a handful. Where it is helpful, the year of publication is shown in brackets so that the reader can more easily get a sense of the way these studies have built upon each other. Table 4 on page 49 lists the important characteristics of each of the studies reviewed below.

It should be noted that in the following sections, the term "failure incidence" is used to denote what Rothman and Greenland[32] refer to as the "incidence proportion", and which is also sometimes called the "unconditional cumulative incidence". It is simply the number of subjects who have experienced sterilisation failure within a specified follow-up time, divided by the total number of subjects. However, as will be noted below, many studies are vague about the duration of follow-up time.

2.5.1 Early case series

Garb (1957) provides an historical literature review of a large number of personal and hospital case series of female surgical sterilisations.[33] Unfortunately no information about the length or nature of the follow-up for any of these case series is reported – thus it is difficult to determine to what degree the estimates of failure incidence are comparable. The reported failure incidences range from 0 to 4.9 per cent. One series, reported by Prystowsky and Eastman, included 1,830 puerperal tubal sterilisations performed at John Hopkins Hospital between 1936 and 1950.[34] Of these, 55 percent were followed up by postal questionnaire. Seventeen failures were discovered (1.7 percent), four of which were ectopic pregnancies. A summary of the case series reviewed by Garb, showing numbers of procedures and failures pooled by type of procedure, is given in Table 1 below.

Table 1 - Historical case series of female sterilisation failure incidence by type of procedure, as reviewed by Garb (1957)

Type of procedure	No. of cases	No. of failures (pregnancies)	Percentage
Madlener	7,829	113	1.44
Pomeroy	5,477	22	0.40
Cornual resection	311	9	2.89
Modified Irving	1,056	0	0
Other or not stated	14,823	66	0.45
All	29,496	210	0.71

McElin, Buckingham and Johnson (1967) reported a hospital case series of 902 sterilisations: 97 percent were Pomeroy procedures, 89 percent were puerperal.[35] Five subsequent pregnancies (0.55 percent) were discovered by opportunistic, non-systematic follow-up. No attempt was made to contact patients or to undertake other forms of active or passive follow-up.

Poulson (1973) reported a hospital case series of 736 tubal sterilisations, with no systematic follow-up.[36] The investigators became aware of six subsequent pregnancies in the series (0.8 percent).

Shah, Courey and Cunanan (1977) described a hospital case series of 3,160 laparoscopic tubal electrocoagulation and division operations.[37] Follow-up was for 6.5 years, although details of follow-up mechanisms are not given. Sixteen pregnancies were recorded (0.5 percent), of which seven were luteal phase pregnancies at the time of the procedure. Nine pregnancies occurred between 12 and 36 months post-procedure (0.28 percent). Hysterosalpingography was performed on the first 150 patients in the case series, which revealed 16 patients (11 percent) with patent tuboperitoneal fistulas – however none of these patients became pregnant and the authors attributed the fistulas to the injection of contrast medium.

Cheng *et al.* (1977) followed up a hospital-based case series for at least two years and found cumulative failure incidences of 0.34, 1.67, 3.12 and 4.49 per 100 women for abdominal (minilaparotomy), culdoscopic, vaginal and laparoscopic procedures respectively.[38] The fact that these cumulative incidences are rather higher than those reported elsewhere in the literature at about the same time may simply be the result of longer and more thorough follow-up (due to the stability and geographical containment of the Singaporean population), or they may be due to the use of life table techniques – this study appears to be the first study of female sterilisation failure to use life tables, despite their widespread and long-standing use in studies of other forms of contraception.. These methodological differences underline the difficulty in comparing reported incidence of failure when follow-up duration and

efficiency are not well described, and when life table techniques are not used to permit known loss to follow-up to be accounted for.

Hughes (1977) reported a hospital case series 1969-76 for Aberdeen, Scotland.[39] The author claimed that this included the vast majority of female sterilisations in north-east Scotland. A total of 9,893 procedures were performed. Follow-up mechanisms and duration were not described. Seventy-seven pregnancies were detected (0.78 percent). Of these, ten were in patients who had open procedures (failure incidence 0.23 percent) and 67 in patients undergoing laparoscopic procedures (failure incidence 1.2 percent). Half of the failures occurred in the first year after surgery.

Keeping, Chang and Morrison (1979) provide a very extensive historical literature review, encompassing some 349 publications covering of all aspects of female surgical sterilisation.[40] Table 2 summarises the incidence of failure in the reports reviewed, classified by timing of procedure as well as type of procedure (note that most of the case series reviewed by Garb and the others reported above are included in this table). It should be remembered that, apart from the extremely variable follow-up periods and procedures used by these studies, that the studies of the laparoscopic procedures were generally performed later than those of the open procedures, and thus may reflect changes in peri-operative arrangements, particularly better screening for luteal phase pregnancies, or limitation of surgery to the pre-ovulatory phase where there is doubt about the reliability of pre-operative contraception.

Table 2 – Incidence of failure of female surgical sterilisation procedures, as reviewed by Keeping, Chang and Morrison (1979)[40]

Type of procedure	Timing of procedure	No. of cases	No. of deaths associated with procedure	No. of pregn-ancies	Incidence of failure (per 100 women)
Open (laparotomy)	Postpartum	10,131	3	41	0.67
Open (laparotomy)	Interval	37,336	2	259	0.71
Laparoscopic	Postpartum/ post-abortion	1,946	0	4	0.27
Laparoscopic	Interval	36,346	2	170	0.47
Vaginal	Either (mostly interval)	3,290	0	8	0.49
All	All	89,049	7	482	0.54

2.5.2 Multi-centre cohorts and nested case-control studies

Mumford, Bhiwandiwala and Chi (1980) assembled data from clinical trials conducted by the International Fertility Research Program in 23 countries, comprising 7,053 laparoscopic Falope-ring procedures, 3,033 minilaparotomy/ Falope-ring procedures and 5,081 minilaparotomy/Pomeroy procedures.[41] The tubal ring cases were also reported separately.[42] All cases were actively followed up for at least one year. The 12 month cumulative incidence of failure, calculated using life-tables, was 0.60 per 100 women for laparoscopic/ring procedures, 0.30 per 100 women for minilaparotomy/Pomeroy procedures and 0.48 per 100 women for minilaparotomy/ring procedures. The incidence of surgical complications was more than twice as great in laparoscopic procedures (2.04 percent) than in those performed by minilaparotomy (0.79 percent).

The authors noted that the main limitation of the study was the fact that it was not an RCT – that is, subjects were not randomised to the procedure types; rather, participating centres offered the types of procedures which they normally performed to potential subjects. Issues of subject recruitment and selection bias were not addressed by the authors. The lack of randomisation of subjects which concerned the authors may well have compromised the validity of comparisons of the incidence of failures and complications for the various sterilisation methods studied, but it has no impact on the validity of their estimates of the incidence of failure of sterilisation technique as they were actually used – the latter being the measure of primary interest to this literature review.

A later study added nine thousand sterilisations by electrocoagulation and spring clip to this cohort. These were also followed up for at least 12 months and analysed in the same manner, with similar results (12-month incidence of pregnancy of less than 1.0 per 100 women).[43]

Chi, Mumford and Gardner (1981) also re-analysed part of the International Fertility Research Program (IFRP) data set in order to investigate differences in incidence of failure for postpartum versus interval laparoscopic sterilisations done using electrocoagulation, Falope-ring and a prototype spring-loaded clip (similar to the Filshie clips in widespread use today).[44] They followed up a cohort of 9,399 women presenting to 33 centres in 19 countries. There were no apparent selection criteria. Approximately 75 percent of this cohort was followed up for at least one year, and apparent luteal phase pregnancies were excluded. Twelve-month life-table incidences of pregnancy were 0.60, 2.64 and 4.35 per 100 women for interval, post-abortion and post-partum procedures respectively (uncontrolled for sterilisation technique). They

posit two plausible reasons for these differences: a) that pregnant women may have higher intrinsic fertility, putting them at higher risk of post-sterilisation failure, and; b) pregnancy may cause physiological changes (such as hyperaemia and oedema) in pelvic structures which make procedural failure more likely. They also found much higher 12 month life-table failure incidence in women sterilised postpartum with spring-loaded clips: 9.31 per 100 women. However, a Mantel-Haenszel stratified analysis of the incidence of failure revealed that gravidity and use of spring-loaded clips were both independent risk factors for sterilisation failure. This appears to have been the first study to undertake such an analysis. A further report based on this dataset was published in 1987, giving similar results.[45]

Chi *et al.* (1980) undertook a nested cumulative case-control study of factors associated with sterilisation failure ("cumulative" here refers to the fact that controls were chosen from non-cases).[46,47] One hundred and sixty cases of non-luteal phase post-procedure pregnancy were detected in a cohort of 14,700 sterilisation procedures which were part of the multi-country IFRP dataset described above. The pregnant cases were matched to non-pregnant controls drawn from the same cohort by surgical centre, anaesthesia type, surgical approach (open, culdoscopic or laparoscopic), tubal occlusion technique and calendar month of sterilisation. A second set of controls was matched by surgical centre, age group (≤ 34 years or 35+ years), parity (≤ 4 or 5+), timing (interval, post-abortion or postpartum), history of pelvic inflammatory disease, history of previous abdominal surgery and lactation status post-sterilisation. Patients who were younger and who were sterilised soon after a new surgical sterilisation programme was established in a given centre were at higher risk of failure, as were patients who did not lactate following the procedure. None of these results are surprising. However, an odds ratio of failure of 7.0 was

found for patients receiving a prototype spring-loaded clip. The authors attributed this to an initial design defect which resulted in inadequate tension on the teeth of the clip when the jaws are closed and locked. Lee and Rubin (1984) also report an investigation of a high incidence of failure (three out of 41 procedures, crude incidence 7.3 per 100 women) in a case series of a private practitioner using Bleier clips. Misapplication of the clips was suspected, but no conclusion was able to be reached.[48]

Ayers *et al*. (1984) investigated nine uterine pregnancies and one ectopic pregnancy in a series of 105 consecutive patients undergoing tubal sterilisation by laparoscopic bipolar electrocoagulation, shortly after bipolar cautery was introduced at their centre.[49] All of the cases of failure became pregnant within three cycles of the procedure. No apparent cause was found. Gunston *et al*. (1983) also reported initially high incidence of failure (4.0 per 100 women at one year of follow-up) with bipolar laparoscopic sterilisation.[50] No further reports of such high incidence of failure with bipolar cautery appear in the literature, making it likely that these reports reflect initial problems adapting to the new, safer, technology.

2.5.3 Other studies from the 1980s and 1990s

Vessey *et al.* (1983) reported on a cohort of 2,243 white British women aged 25-39 years, currently using contraception, who had undergone mainly laparoscopic tubal sterilisation in 17 UK clinics between 1968 and 1981.[51] It is unclear whether there was any self-selection of the cohort. Extensive and active follow-up was performed for at least 12 months, and in some subjects, for up to 8 years. The overall incidence of failure in the first 12 months following surgery was 0.37 per 100 women, and 0.10

per 100 women subsequently. A life-table analysis revealed cumulative incidences of failure of 0.4 per 100 women after 1 year, 0.8 after 4 years and 1.0 after 7 years.

Sitompul *et al*. (1984) undertook a randomised trial of three surgical approaches to tubal sterilisation in 300 women presenting to a university hospital.[52] Follow-up to four years was completed for 295 subjects. Only one post-operative pregnancy was discovered, three years post-operatively.

Sherman and Burigo (1984) reported a non-randomised trial of laparoscopic versus minilaparotomy Falope-ring sterilisation in which active follow-up at one year revealed 2 failures in 335 cases (0.6 per 100 women).[53]

Aranda *et al*. (1985) conducted a well-designed randomised trial of Yoon band versus Rocket clip sterilisations via minilaparotomy in 663 successive patients at three centres.[54] The proportion of subjects attending follow-up clinics at six, twelve and 24 months were 85, 61 and 47 percent respectively. These figures provide some indication of the completeness of follow-up when passive methods such as relying on patients returning to a follow-up clinic are used. Analysis was by life-table, with censoring at the date of the last clinic follow-up. Three pregnancies were detected in each group, giving 24 month cumulative life-table incidence of failure of 1.0 and 0.9 per 100 women in the ring and clip groups respectively.

The Indian Council for Medical Research (1984) reported an early (within six months) failure incidence of 6.9 per 100 women in a case series of 869 postpartum or post-abortion Filshie clip sterilisations, shortly after the use of this device was introduced into India.[55] De Villiers (1987) reported a postpartum 12-month failure

incidence of 1.0 per 100 women with the same device shortly after its use began in South Africa.[56]

Chi *et al*. (1991) undertook a prospective cohort study of the timing of postpartum sterilisation using Filshie clips applied via minilaparotomy – those sterilised within 48 hours of delivery were compared to those sterilised later. Life-table incidence of pregnancy at 12 and 24 months was similar in both groups.[57]

De Villiers (1992) later compared five methods for postpartum sterilisation.[58] Details of duration of follow-up were not given in English (the body of the paper is in Afrikaans) and hence the incidences of failure of between 0.5 and 2.1 per 100 women are difficult to interpret. However, the author notes that female surgical sterilisation is far from perfect and pleads for the legalisation of abortion after failed sterilisation in his country.

Yan *et al*. (1990) carried out a prospective randomised trial of Filshie clips versus the Pomeroy procedure in 200 postpartum women.[59] Follow-up at clinic attendances at 6, 12 and 48 months was carried out. Although over 95 percent of subjects attended at least one follow-up clinic, many did not attend them all but the authors did not give details. Life-table analysis with censoring was not undertaken. One failure in the Pomeroy group was found at 6 months, and none in the Filshie clip group.

Stovall *et al.* (1991) undertook a similar randomised trial of Falope-ring versus Hulka-Clemens clips, with active follow-up to 16 months.[60] Cumulative failure incidences were eight in 176 women (4.5 percent) in the clip group, and five in 189 (2.6 percent) in the ring group.

Birdsall *et al*. (1994) retrospectively reviewed all female sterilisation procedures, the majority by laparoscopic application of Filshie clips, at the National Womens' Hospital, Auckland in 1988 and 1989.[61] They undertook extensive active follow-up by post, telephone and through general practitioners, but unfortunately failed to record the exact timing and length of follow-up. However, it may have been as long as four or five years post-procedure, given that their paper was published five years after the cohort was operated upon. They reported a cumulative failure incidence of 1.4 per 100 women, excluding 7 luteal phase pregnancies (0.65 per 100 women). They note that these incidences of failure are at least three times higher than the 0.5 percent quoted to patients, based on overseas studies. Their higher incidence may have been due to their longer period of follow-up compared to the majority of previous studies, and/or the use of life-table techniques in their analysis (although they did not explicitly state that life-table analysis was used).

Makar *et al*. (1990) reported a retrospective study of 1,437 sterilisations in a hospital in Belgium.[62] A failure "rate" of 1.18 per cent was mentioned in the abstract, but no details of the study design, follow-up procedures or duration or the method of calculation for this result appeared in the body of the paper.

Trias *et al*. reported on a very large study of nearly 45,000 tubal sterilisations carried out between 1973 and 1982 in a large clinic in Bogota, Columbia.[63] This study appears to have included all women undergoing surgical sterilisation procedures carried out at that clinic. Ninety four percent of the procedures were laparoscopic, using unipolar electrocoagulation or Yoon Silastic band, and all were interval procedures. Unfortunately, no active follow-up was undertaken – the study relied on

failures being reported to the clinic, either directly by patients, or indirectly by third parties. Subjects were right censored 60 months after their procedure – that is, follow-up was assumed to be complete for a five year period. Sixty one luteal phase pregnancies were excluded, and eight failures occurring after 60 months were also excluded due to censoring, leaving some 503 cases in the denominators of the life table analysis. The results are summarised in Table 3 below.

Table 3 - Trias *et al.* (1987)[63]: Cumulative life table incidence of pregnancy following tubal sterilisation

Method/Age group	Period	Failures per 100 women, by months since procedure (95% CI in parentheses)				
		12	24	36	48	60
All	1973-82	0.6 (0.6-0.7)	1.0 (0.9-1.1)	1.3 (1.2-1.4)	1.4 (1.3-1.5)	1.4 (1.3-1.6)
Minilaparotomy/ Pomeroy	1974-82	0.1 (0.0-0.2)	0.4 (0.2-0.7)	0.7 (0.4-1.1)	0.7 (0.4-1.1)	0.9 (0.5-1.2)
Laparoscopy	1973-82	0.7 (0.6-0.8)	1.1 (1.0-1.2)	1.3 (1.2-1.5)	1.4 (1.3-1.5)	1.5 (1.3-1.6)
Laparoscopy/ Cautery	1973-82	0.3 (0.1-0.5)	0.4 (0.2-0.6)	0.5 (0.3-0.7)	0.5 (0.3-0.8)	0.6 (0.3-0.8)
Laparoscopy/ Bands	1973-75	0.7 (0.6-0.8)	1.1 (0.9-1.2)	1.3 (1.2-1.5)	1.4 (1.3-1.6)	1.5 (1.4-1.7)
	1981-82	1.0 (0.8-1.2)	2.1 (1.7-2.5)	-	-	-
Age < 32 years	1981-82	1.3 (1.0-1.7)	2.6 (2.0-3.1)	-	-	
Age 32+ years	1981-82	0.7 (0.4-0.9)	1.6 (1.2-2.1)	-	-	-

This study illustrates the same methodological issue facing the current study. The lack of active follow-up tends to cause the numerator for the failure incidence to be underestimated. However, the length of effective follow-up for each subject is not truly known, but rather is assumed to be for some arbitrary period (in the case of the

Columbian study), or for the duration of available data to be linked (as in this study). Thus, the denominator will tend to be overestimated, because some subjects will not be censored as early as they should be. Hence, failure incidence estimated by such studies will always be lower than the true incidence, assuming the absence of egregious methodological flaws which might cause over-estimation of failure incidence. In the context of this thesis, an example of such a flaw might be an excessive false linkage rate, such that a large number of records for women undergoing surgical sterilisation are incorrectly linked to subsequent records indicating conception. The implication is that given the methodological limitation of only being able to estimate a lower bound of the incidence of failure (as discussed above), it is important that record linkage is performed in a "conservative" manner that minimises the number of false links.

Lassner *et al*. (1988) conducted a study of 13,423 tubal sterilisation procedures performed in Rio de Janeiro between 1981 and 1984.[64] The study design was very similar to that used in the Colombian study. They found life-table cumulative incidences of failure of 0.54 per 100 sterilisations at 12 months, and 1.04 per 100 at 48 months. However, unlike the Columbian study, they undertook active follow-up on a random 15 percent sample of their cohort in order to estimate the efficiency of the passive follow-up used for the rest of the cohort. The results of this validation are not reported in any detail, but they assert that no additional pregnancies were detected as a result of active follow-up, which led them to consider their passive follow-up to be reasonably complete.

Peterson *et al*. (1996) reported the results of the US Collaborative Review of Sterilization study (CREST).[65] This was a prospective study of women undergoing

tubal sterilisation between 1978 and 1986 in one of 14 medical centres in 14 different states of the US. Eligibility for entry into the study was restricted to women who underwent interval laparoscopic procedures (band or clip) and to women undergoing interval or postpartum open procedures. Women were not randomised to particular types of procedure. No information is given on the number or characteristics of women who did not consent to be entered into the study. Study subjects were followed-up by telephone one month after their procedure and then yearly for at least five years, by the same method. Women enrolled in the earlier years of the study were followed up in this fashion for up to 14 years. If follow-up contact was lost, the subject was censored at the date of the last successful follow-up. At follow-up, women were asked if they had become pregnant – if so, confirmation through medical records was sought. Apart from pregnancy, subjects were censored if they had a repeat sterilisation procedure, a tubal anastomosis or other reversal attempt, a hysterectomy, if they died, or if they refused to be interviewed further. Luteal phase pregnancies were excluded from the analysis, and if there was doubt whether a failure was a luteal phase pregnancy, it was excluded regardless. Of the 10,863 women enrolled, 178 were excluded for the reasons mentioned above (116 due to loss to follow-up). Analysis was by life-table and Cox proportional hazards regression. One hundred and forty-three women had pregnancies classified as true failure, resulting in cumulative life-table incidences of failure per 100 procedures, at one, two, four and ten years respectively of 0.55, 0.84, 1.18 and 1.85.

These incidences are comparable to those of the Columbian study. One new methodological issue was raised by the US study: should self-reported but otherwise unverified spontaneous abortions be counted as failures? When self-reported, unverified spontaneous abortions were excluded, the 10-year cumulative failure

incidence dropped from 1.85 per 100 procedures to 1.66 per 100 procedures. The corollary for record linkage studies of sterilisation failure is that admissions to hospital for dilatation and curettage (D&C) procedures, where a diagnosis code indicating spontaneous (or induced) abortion is also present, should be counted as a failure, but other admissions for D&C procedures (which are very common) should not.

2.5.4 Summary of studies reporting incidence of failure

Table 4, which begins on the next page, lists the main attributes of the studies reviewed above. All of the studies have a similar design: a cohort of patients who have undergone surgical sterilisation procedures are followed over time to determine whether they conceive. The factors which distinguish the study designs include: a) the cohort selection or recruitment mechanism; b) whether patients were randomised to undergo particular types of procedures; c) whether data were collected prospectively or retrospectively; d) the nature and duration of follow-up mechanisms; and e) whether life-table techniques, incorporating censoring information, were used to calculate the failure incidence.

The term “hospital case series/cohort” has been used for the most common design, in which the cohort being followed-up, usually retrospectively, comprises all (or most) patients undergoing surgical sterilisation at a particular hospital during a particular period of time. A “personal case series/cohort” is similar but includes subjects operated on by only one surgeon. A “Pooled multi-centre case series/cohort” includes subjects operated on at multiple hospitals or clinics, but analysed together. The term “clinical trial” has been used to denote a study using prospective data

collection, usually with the intent of comparing the incidence of failure or complications between different procedures or operative timings or approaches.

Table 4 - Studies reporting incidence of failure of female surgical sterilisation

Author (yr)	Ref. No.	Study type	Procedure type(s) and timing	No. of subjects	Follow-up duration	Reported failure incidence	Comments
Garb (1957)	33	Review of multiple case-series/cohorts	Various	29,496	Not stated	0.71 per 100 women	Pooled results from multiple case series - see Table 1 on page 34 for more details
Prystowsky and Eastman (1955)	34	Hospital case series/cohort	Various, all postpartum	1,830	Postal questionnaire after 2-4 years with 55% response rate	1.7 per 100 women	Four of 17 failures were ectopic pregnancies.
McElin, Buckingham and Johnson (1967)	35	Hospital case series/cohort	Pomeroy, 89% postpartum	920	Opportunistic passive follow-up of unspecified duration through subsequent encounters at same clinic, no active recall or tracing of patients who did not attend.	0.55 per 100 women	

Author (yr)	Ref. No.	Study type	Procedure type(s) and timing	No. of subjects	Follow-up duration	Reported failure incidence	Comments
Poulson (1973)	36	Hospital case series/cohort	Various	736	Opportunistic passive follow-up of unspecified duration through subsequent encounters at same clinic, no active recall or tracing of patients who did not attend.	0.8 per 100 women	
Shah, Courey and Cunanan (1977)	37	Hospital case series/cohort	Laparoscopic electrocoagulation	3,160	6.5 years of follow-up, methods not given	0.5 per 100 women	Seven of 16 pregnancies thought to be luteal phase, remaining failure incidence 0.28 per 100 women
Cheng *et al.* (1977)	38	Hospital case series/cohort	(a) Minilaparotomy, (b) culdoscopic, (c) vaginal (d) laparoscopic approaches	2,156	At least two years of systematic active follow-up by telephone and mail	(a) 0.34, (b) 1.67, (c) 3.12, (d) 4.49 per 100 women	
Hughes (1977)	39	Hospital case series/cohort	Various	9,893	Not described	0.78 per 100 women	

Author (yr)	Ref. No.	Study type	Procedure type(s) and timing	No. of subjects	Follow-up duration	Reported failure incidence	Comments
Keeping, Chang and Morrison (1979)	40	Pooled multi-centre case-series/cohorts	Various	89,049	Various	0.54 per 100 women	See Table 2 on page 34 for more details.
Mumford, Bhiwandiwala and Chi (1980)	41	Pooled unrandomised multi-centre clinical trials	(a) laparoscopic Falope-ring procedures, (b) minilaparotomy/ Falope-ring procedures, (c) minilaparotomy/P omeroy	(a) 7,053 (b) 3,033 (c) 5,081	Active follow-up for at least one year by telephone and/or mail.	(a) 0.60 , (b) 0.48, (c) 0.30 per 100 women	
Bhiwandiwala, Mumford and Feldblum (1982)	42	Pooled unrandomised multi-centre clinical trials	Electro-coagulation and spring clip	9,123	Active follow-up for at least one year by telephone and/or mail	< 1.0 per 100 women	

Author (yr)	Ref. No.	Study type	Procedure type(s) and timing	No. of subjects	Follow-up duration	Reported failure incidence	Comments
Chi, Mumford and Gardner (1981) and Chi *et al.* (1987)	44, 45	Pooled unrandomised multi-centre clinical trials	Electro-coagulation, Falope-ring and a prototype spring-loaded clip	9,399	Active follow-up for at least one year by telephone and/or mail	0.60, 2.64 and 4.35 per 100 women for interval, post-abortion and post-partum procedures respectively	9.31 failures per 100 women found with spring clips applied post-partum.
Ayers *et al*. (1984)	49	Personal case-series/cohort	Laparoscopic bipolar electrocoagulation	105	Three months active follow-up by telephone/GP	9.5 per 100 women	Soon after introduction of bipolar cautery to this clinic
Gunston *et al*. (1983)	50	Hospital case series/cohort	Various	9,430	Opportunistic passive follow-up of unspecified duration through subsequent encounters at same clinic, no active recall or tracing of patients who did not attend.	0.25 per 100 women	4.0 failures per 100 women for bipolar cautery.

Author (yr)	Ref. No.	Study type	Procedure type(s) and timing	No. of subjects	Follow-up duration	Reported failure incidence	Comments
Vessey *et al.* (1983)	51	Pooled multi-centre case-series/cohorts	Laparoscopic approach	2,243	Active follow-up for one to eight years by mail or tracing via GPs	0.4 per 100 women after 1 year, 0.8 after 4 years and 1.0 after 7 years	Life-table cumulative incidence of failure
Sitompul *et al*. (1984)	52	Randomised clinical trial	Randomised between laparoscopy, culdoscopy and minilaparotomy approaches	300	Active follow-up by telephone or mail for four years	0.33 per 100 women after 4 years follow-up	
Sherman and Burigo (1984)	53	Non-randomised clinical trial	Laparoscopic versus minilaparotomy Falope-ring sterilisation	335	Active follow-up for one year, mechanism not stated	0.6 per 100 women	

Author (yr)	Ref. No.	Study type	Procedure type(s) and timing	No. of subjects	Follow-up duration	Reported failure incidence	Comments
Aranda *et al*. (1985)	54	Randomised clinical trial	Yoon band versus Rocket clip sterilisations via minilaparotomy	663	Opportunistic passive follow-up for up to 24 months through subsequent encounters at same clinic, no active recall or tracing of patients who did not attend.	1.0 and 0.9 per 100 women after 24 months in the ring and clip groups respectively	Cumulative life-table incidence of failure. Proportion of cohort attending follow-up clinics: 85, 61 and 47 percent at six, 12 and 24 months respectively
Indian Council for Medical Research (1984)	55	Hospital case series/cohort	Post-partum Filshie clips	869	Opportunistic passive follow-up for 6 months through subsequent encounters at same clinic, no active recall or tracing of patients who did not attend.	6.9 per 100 women after 6 months	Study performed immediately after the introduction of Filshie clips into India

Author (yr)	Ref. No.	Study type	Procedure type(s) and timing	No. of subjects	Follow-up duration	Reported failure incidence	Comments
De Villiers (1987)	56	Hospital case series/cohort	Filshie clips	789	Opportunistic passive follow-up of unspecified duration through subsequent encounters at same clinic, no active recall or tracing of patients who did not attend.	1.0 per 100 women	
De Villiers (1992)	58	Hospital case series/cohort	(a) Vienna (b) Pomeroy (c) total fimbriectomy (d) Filshie clip (e) Irving	(a) 3580 (b) 892 (c) 1578 (d) 808 (e) 456	Unknown (body of paper is in Afrikaans), probably similar to de Villiers (1987)	(a) 0.5 (b) 2.1 (c) 1.5 (d) 2.2 (e) 0.2 per 100 women	
Yan *et al*. (1990)	59	Randomised clinical trial	Filshie clips versus Pomeroy	200	Opportunistic passive follow-up through subsequent encounters at same clinic at 6, 12 and 48 months, no active recall or tracing of patients who did not attend.	< 1.0 per 100 women in Pomeroy group at 6 months, no known failures in Filshie clip group	

Author (yr)	Ref. No.	Study type	Procedure type(s) and timing	No. of subjects	Follow-up duration	Reported failure incidence	Comments
Stovall *et al.* (1991)	60	Randomised clinical trial	(a) Falope-ring versus (b) Hulka-Clemens clip	189 (ring), 176 (clip)	Active follow-up to 16 months post-procedure, methods not stated	(a) 2.6, (b) 4.5 per 100 women at 16 months	Cumulative life-table incidence of failure
Birdsall *et al*. (1994)	61	Hospital case series/cohort	Majority by laparoscopic application of Filshie clip	1094	Active follow-up on a single occasion 5-10 years after the procedure by telephone, post and contact tracing via GPs	1.4 per 100 women (1.2 for Filshie clips)	Unclear whether life-table techniques were not used in the analysis.
Makar *et al*. (1990)	62	Hospital case-series/cohort	Bipolar electrocoagulation	1437	Not stated, but study done 6 years after procedures performed	1.18 per 100 women	
Trias *et al*. (1987)	63	Hospital case series/cohort	Unipolar electrocoagulation or Yoon band, laparoscopic interval procedures	Approx. 45,000	Passive follow-up for 60 months through clinic attendances and third-party reports	0.6, 1.0, 1.3, 1.4, 1.4 per 100 women at 12, 24,36,48 and 60 months	Cumulative lifetable incidence of failure. Life-table analysis censored at 60 months, luteal phase pregnancies excluded. See Table 3 on page 44 for more details of results

Author (yr)	Ref. No.	Study type	Procedure type(s) and timing	No. of subjects	Follow-up duration	Reported failure incidence	Comments
Lassner *et al*. (1988)	64	Multi-centre case series/cohort	98% by Yoon band applied laparoscopically	13,423	Passive follow-up for 48 months through clinic attendances, with active follow-up by post., telephone and contact tracing of a 15% sub-sample	0.54, 1.04 per 100 women at 12 and 48 months resp.	Cumulative lifetable incidence of failure. No significant difference in pregnancy ascertainment was found between passive follow-and active follow-up groups.
Peterson *et al*. (1996)	65	Multi-centre non-randomised clinical trial	Laparoscopic band or clip.	10,685	Active follow-up by annual telephone calls for up to 14 years	0.55, 0.84, 1.18 and 1.85 per 100 women at 12, 24,48 and 120 months resp.	Cumulative lifetable incidence of failure.

2.5.5 Studies of luteal phase pregnancies

Luteal phase pregnancies are mentioned by many authors as a real or potential methodological problem for their studies of female sterilisation failure. Two studies have examined this issue in detail.

Loffer and Pent (1980) reported six luteal phase pregnancies in 2,249 sterilisations (0.27 percent) in which no precautions were taken to exclude pre-operative pregnancy.[66] This is similar to the incidence of luteal phase pregnancy reported by Hulka.[67] None of the women in the Loffer and Pent series were offered a second sterilisation procedure, and none became pregnant subsequently. They also report results from an early series of 24,764 laparoscopic cases in which there were 60 luteal phase pregnancies (0.24 percent) and 44 other failures (0.17 percent).[68] Chi and Feinblum (1981) undertook a case-control study in which 37 luteal phase pregnancies (LPP) were matched with 74 non-pregnant controls, and with 123 "true" sterilisation failures (TSF).[69] They found that women with LPP were more fecund, as measured by the proportion which had last been pregnant less than two years prior to the sterilisation procedure. There were no other significant differences, except that LPPs were less likely to be terminated than TSFs.

The improved sensitivity of modern radio-immunoassays (RIA) tests for beta-hCG may have substantially reduced the problem of surgical sterilisation being performed on women who have just conceived, provided that sterilisation candidates are in fact screened prior to surgery.

2.6 Ectopic pregnancy following tubal sterilisation

Tubal sterilisation is well known to be a risk factor for ectopic pregnancy, due to the 30- to 50-fold difference in the cross-sectional size of spermatazoa and ova: micro-fistulae formation or re-cannulisation of the proximal tubal stumps are far more likely to let spermatozoa into the peritoneal cavity than they are to admit ova into the uterus. There is a very large literature on ectopic pregnancy. However, a manageable number of papers dealing primarily with the subject of ectopic pregnancy after tubal sterilisation were located, and these will be reviewed very briefly here.

Seventeen of the papers are essentially clinical case series (or anecdotal) reviews of ectopic pregnancies occurring in women who had previously undergone tubal sterilisation, and will not be covered in any detail here.[70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86]

Apart from discussion of clinical characteristics and anatomical pathology, these papers tend to report the proportion of ectopic pregnancies in the case series in question which were subsequent to tubal sterilisation. Values range from 0.6 percent to 67 percent.[86,73] McCausland undertook an extensive review of the literature on this subject in 1980 and calculated that a weighted mean of 12.3 percent of ectopic pregnancies reported in the literature were subsequent to tubal sterilisation.[77]

Six papers reporting on studies which attempted to quantify the association between tubal sterilisation and risk of ectopic pregnancy were located. The main attributes of these studies are listed in Table 5 below.

Honore and O'Hara (1978) used an estimate from the literature of the overall incidence of failure of tubal sterilisation (provided by Garb) to calculate the expected number of post-sterilisation pregnancies in their population (in Newfoundland, Canada) over a seven year period.[87,33] They then applied the observed probability of ectopic pregnancy in that population (0.27 percent of all pregnancies) to the expected number of post-sterilisation pregnancies to arrive at an expected number of ectopic pregnancies following sterilisation, assuming that the risk of ectopic pregnancy is independent of previous tubal sterilisation. Comparing this with the observed number of post-sterilisation ectopic pregnancies gives a rough estimate of the incidence density ratio (IDR) of ectopic pregnancy following sterilisation compared to no sterilisation: an IDR of 20 was reported. These calculations are similar to those used in calculating indirectly standardised mortality ratios, except that no age stratification was used. A theoretical criticism is that a significant proportion of all ectopic pregnancies used to calculate the expected number post-sterilisation were themselves post-sterilisation, thus biasing the estimated risk ratio towards zero (and not towards the null, or 1.0). McCausland's estimate of the attributable fraction of ectopic pregnancies due to surgical sterilisation could have been used to correct the expected number.

DeStefano *et al*. (1982) correctly pointed out that although the relative risk of ectopic pregnancy (versus intrauterine pregnancy) in women who had undergone tubal sterilisation (compared to women who had not) was significantly greater than one, it did not follow that the absolute risk of ectopic pregnancy was higher, as had been hypothesised by some authors.[88,76,81] The authors used the cumulative annual post-sterilisation failure incidences reported by Cheng *et al.* and a conservative estimate derived from the literature that about one-sixth of those failures will be ectopic

pregnancies.[38] Annual incidence rates of ectopic pregnancy in women using other forms or contraception, or no contraception, were derived from studies by Vessey *et al*., Ory and Westrom *et al*.[89,90,91] Some form of life-table calculations (not exactly specified) were then used to calculate the cumulative risk of an ectopic pregnancy through to age 45 years. They illustrated that the cumulative risk of ectopic pregnancy for women undergoing tubal sterilisation was 20 to 40 times lower than for women using no contraception, and 3 to 10 times lower than for women using an IUD, but slightly higher than for women using barrier contraception, and at least an order of magnitude higher than for women using the oral contraceptive pill.

The World Health Organisation Task Force on Intrauterine Devices for Fertility Regulation (1985) conducted a multinational case-control study in which 1108 cases of ectopic pregnancy were matched against pregnant and non-pregnant controls.[92] When cases were compared to pregnant controls there were statistically significantly increased odds ratios (OR) for ectopic pregnancy associated with the use of an intrauterine device (OR = 6.4) and tubal sterilisation (OR = 10.9). When compared to non-pregnant controls, the relative risk of ectopic pregnancy was reduced by all methods of contraception, with odds ratios of 0.5 for the IUD, 0.1 for the oral contraceptive pill, and 0.2 for tubal sterilisation. These results are consistent with the estimated relative risks of ectopic pregnancy calculated by DeStefano *et al.* several years earlier.[88]

Marchbanks *et al.* (1988) conducted a case-control study of all 274 cases of ectopic pregnancy in the town of Rochester, Minnesota, occurring between 1935 and 1982, and 548 matched controls selected from women delivered of live-born infants.[93] Twenty-two potential risk factors were examined. Twelve of these were entered into

a conditional logistic regression model, which revealed four risk factors with statistically significant odds ratios: current IUD use (adjusted OR 13.7, 95% CI 1.6-120.6), history of infertility (adjusted OR 2.6, 95% CI 1.6-4.2), pelvic inflammatory disease (adjusted OR 3.3, 95% CI 1.6-6.6) and previous tubal surgery (adjusted OR 4.5, 95% CI 1.5-13.9).

Holt *et al*. (1991) matched 249 cases of ectopic pregnancy occurring in members of a US health maintenance organisation between 1981 and 1986 with 835 randomly selected controls (also members of the same HMO), matched on age and county of residence.[94] After controlling for various potential confounders, the odds ratio for previous tubal sterilisation performed as an interval procedure were 3.7 (95% CI 1.7-8.0) compared to oral contraceptives, 2.8 (1.5-5.5) compared to barrier methods and 0.2 (0.1-03) when compared to no contraception at all. Odds ratios were considerably lower for tubal sterilisations performed post-partum (1.2, 0.9 and 0.1 respectively).

Peterson *et al*. (1997) examined risk factors for ectopic pregnancy in the multi-centre cohort of women who had undergone tubal sterilisation which they reported on in their earlier paper. [95,65] They found that the cumulative probability (that is, the absolute risk) of ectopic pregnancy after all forms of tubal sterilisation was 0.07, 0.4 and 0.73 per 100 procedures after one, five and ten years respectively. However, the hazard varied considerably between types of procedure, with the hazard following bipolar electrocoagulation in particular some two to three times higher than for other methods. Furthermore, the hazard appeared to increase four or more years after the procedure, whereas it appeared to decline after four or five years for the other techniques. A history of pelvic inflammatory disease, black race and being under 30 years of age were all independent risk factors.

Table 5 - Studies on the risk of ectopic pregnancy following tubal sterilisation in females

Authors	Ref No.	Year	Study type	No. of subjects	Results	Comments
Honore and O'Hara	87	1978	Estimation of population incidence	N/A	IDR of ectopic pregnancy after sterilisation compared to entire population ~ 20	Calculations based on historical incidence of sterilisation failure
DeStefano *et al.*	88	1982	Relative survival analysis	N/A	Cumulative HR to 45 yrs of ectopic pregnancy after sterilisation < 0.05 (relative to no contraception)	HR > 10 relative to women using OCP.
WHO	92	1985	Multicentre case-control	1108 cases	OR for ectopic after sterilisation = 10.9 with pregnant controls, OR=0.2 with non-pregnant controls.	OR for ectopic with IUD = 6.4 with pregnant controls, 0.5 with non-pregnant controls. All ORs statistically significant.
Marchbanks *et al.*	93	1988	Case-control	274 cases	Adjusted OR for ectopic in women who had undergone any form of tubal surgery=4.5 (95% CI 1.5-13.9)	Did not distinguish between types of tubal surgery
Holt *et al.*	94	1991	Matched case-control	249 cases, 835 controls	Adjusted OR=3.7 (95% CI 1.7-8.0) for interval procedures, OCP as baseline.	OR for tubal sterilisation performed post-partum=1.2
Peterson *et al.*	95	1997	Cohort	10,685 followed -up	Cumulative (absolute) risk of ectopic after tubal sterilisation 0.07, 0.4 and 0.73 per 100 women at 1, 5 and 10 years respectively	Hazard declines four or five years post-procedure except if bipolar coagulation used.

Finally, mentioned for the sake of completeness, Hendrix *et al*. (1998) undertook a study in which they selected 208 consecutive cases of ectopic pregnancy in women who had previously undergone tubal sterilisation and who were admitted (with ectopic pregnancy) to two US hospitals between 1992 and 1995.[96] These were compared with an identical number of ectopic pregnancy presentations in women who had no history of tubal surgery – the next suitable case of ectopic pregnancy was chosen for this purpose. The authors describe this as a case-control study, although it clearly is not (nor is it a cohort study, or any variation of such) – indeed, it is difficult to understand the epidemiological basis for the comparisons they performed.

2.7 Risk of hysterectomy and death following tubal sterilisation

The literature on these topics will be reviewed very briefly (and non-exhaustively), as although they are somewhat peripheral to the subject of this thesis, both hysterectomy and death are important censoring events in any study of tubal sterilisation failure.

Templeton and Cole (1982) used the Scottish maternity and General Hospital discharge summary database to assemble a cohort of patients aged 30 to 39 years who underwent tubal sterilisation in 1973.[97] This cohort was then linked to gynaecological discharge summaries for the four subsequent years using the Scottish record linkage system.[98] A control group, comprising women in the same age group who were delivered of a live infant or stillbirth in 1973 and were not subsequently sterilised or pregnant, was linked to gynaecological discharge summaries in the same way. The authors found that the relative risk for hysterectomy in the sterilised group was 3.7, even after age adjustment. Unfortunately, the authors did not adjust for other

possible confounders such as socio-economic class, despite having this information available to them.

Cohen (1987) used a similar study design with person-level health insurance clams for all of Manitoba.[99] Follow-up for eight years by record linkage of a cohort of 4,374 tubal sterilisation patients and 6,835 comparison subject (a random population-based sample) was achieved. Cox proportional hazards regression was used to adjust for possible confounders of previous gynaecological history, marital status, number of physician visits and number of hospitalisations. A hazard ratio for hysterectomy of 1.6 (95% CI 1.2-2.3) was found for sterilised women aged 25 to 29 years, but only 1.1 (95% CI 0.8-1.3) for sterilised women aged 30 to 44 years.

Goldhaber *et al*. (1993) conducted an even larger study of similar design, using record linkage to follow up more than 80,000 women who were members of a large US health maintenance organisation.[100] Sterilised women were more likely to undergo subsequent hysterectomy (RR=1.35, 95% CI 1.26-1.44), with a higher risk in the younger 20-24 year age group (RR=2.45), declining to one in the oldest age group.

These three studies are significant in that they are the only instances encountered in this review in which record linkage was used — in the first two on a population basis — to investigate outcomes after tubal sterilisation. However, no published studies were discovered which used population-based record linkage to estimate the incidence of failure of tubal sterilisation.

Studies of similar design, but not population-based or using record linkage for follow-up, by Stergachis *et al.* (1990) and Hillis *et al*. (1998) reported similar risks for hysterectomy following tubal sterilisation.[101,102]

There have been several studies of disturbance of menstrual function following tubal sterilisation, with inconclusive results – some studies found virtually no change, others reported increased incidence of menorrhagia and DUB (dysfunctional uterine bleeding).[103, 104, 105, 106, 107, 108, 109, 110, 111] Some authors have suggested that such menstrual disturbances may be responsible for the increased risk of hysterectomy following tubal sterilisation, although no study has demonstrated a causal link or association.

Several studies have examined mortality associated with sterilisation.[112,113,114,115,116] Although peri-operative mortality of between 3 and 53 deaths per 100,000 procedures was reported, none of these studies found any increase in long-term mortality associated with tubal sterilisation, which is important information for long-term follow-up studies. The relevance for this study is that women of reproductive age who have undergone surgical sterilisation can be expected to experience the same rather low death rates (between 32.4 and 69.7 per 100,000 per annum in 2003 for Australian women in age groups 20-24 through to 35-39 years[117]) as other women of similar age. Hence, although death is obviously a censoring event, the failure to completely ascertain all deaths in the cohort being followed in this study is unlikely to have a material impact on the results.

Rubin *et al*. (1982) noted that one-year mortality risks for various methods of contraception in developed countries were 0.1 per 100,000 procedures for

vasectomies, 0.3 for IUD use, 2.2 for legal abortion, 4.0 for female surgical sterilisation, and 18.7 for pregnancy and delivery.[118] Ten-year cumulative risk was identical for the once-only procedures, but increased to 3.0 per 100,000 for IUD use and was 12 per 100,000 for the lowest risk category of oral contraceptive pill (OCP) users, and more for higher risk OCP users.

2.8 Regret and reversal

Westhoff and Davis[119] (2000) provide a good review of the US literature on reversal success rates: reported incidence of term pregnancy following reversal varies from 42 to 52 percent for unipolar or bipolar cautery sterilisations,[120, 121, 122] 41 to 74 percent for Pomeroy procedures,[121, 122, 123] and 83 to 87 percent for ring and Hulka clip procedures.[120, 121, 124] These success rates dictate that any indication of attempted reversal be treated as a censoring event in follow-up studies of sterilisation effectiveness.

No literature on population-based rates for reversal procedures was located. However, demand for reversal procedures is likely to be driven by regret. Grubb *et al*. (1985) used early data from the Collaborative Review of Sterilization study to obtain estimates that 2.0 percent of subjects regretted having had a tubal sterilization at one year after the procedure and 2.7% after two years.[125] Hillis *et al*. repeated the analysis on a final version of the same data set and found cumulative probabilities of expressed regret at up to 14 years after sterilisation of 20.3 percent in women aged 30 years or younger at the time of sterilisation, and 5.9 percent in older women.[126]

None of the studies of sterilisation failure reviewed mention the use of assisted conception technologies such as IVF (in-vitro fertilisation) or TUG (trans-uterine

GIFT [gamete intra-fallopian transfer]) as possible censoring events, probably because these techniques were unknown or uncommon when the studies were conducted.

2.9 Other related literature

There is a large literature on operative and peri-operative costs and complications of tubal sterilisation procedures which has not been reviewed for this thesis.

The Cochrane Collaboration has recently published two extensive reviews and meta-analyses relating to tubal sterilisation. The first, by Nardin *et al*. (2003) undertook a meta-analysis of published studies in order to compare major and minor morbidity, incidence of failure and acceptability (for both the subjects and the surgeons).[127] Only randomised controlled trials which compared different techniques for tubal sterilisation, regardless of the surgical approach (minilaparotomy, culdotomy or laparoscopy) were considered. With respect to failure incidence, no statistically significant difference was found between tubal rings versus clips, based on pooled results for three studies (Peto OR 0.70, 95% CI 0.28-1.76). Only one pregnancy occurred in the two randomised studies of modified Pomeroy operations versus electrocoagulation – in other words, any comparison of failure incidences would have extremely low power, due to the small numbers involved. The second, by Kulier *et al*. (2004) evaluated differences in operative and peri-operative incidence of complications between minilaparotomy and endoscopic approaches to tubal sterilisation, and is not directly relevant to this thesis.[128]

Argent (1985) reviewed UK case law regarding failed female surgical sterilisation, and noted that patients need to be informed of the possibility and risk of failure using the best available data for the jurisdiction in question.[129]

2.10 Summary

The findings of this literature review, and its implications for the current study can be summarised as follows:

- There are large numbers of earlier studies which report incidences of failure of 1 per 100 procedures or less. However, most of these earlier studies undertook only passive follow-up for less than two years, on the assumption that failures were most likely to occur in the first year after the procedure.
- Studies which used active follow-up for periods of greater than one year have tended to find longer-term incidences of failure of between 1 and 2 per 100 women.
- Most of the earlier studies failed to use life-table techniques in the estimation of failure incidence, thereby including women who were not effectively being followed-up in the denominator, causing further underestimation of the true incidence of failure.
- Luteal phase pregnancies are a significant cause of apparent failure of surgical sterilisation, and care must be taken to exclude them if the "technical failure rate" of a method is to be properly assessed. It is also good practice for clinicians to screen candidates for surgical sterilisation for current pregnancy, immediately before the procedure.
- Many studies found higher incidences of failure for postpartum and post-abortion procedures, compared to interval procedures.

- There has only been one large study in which active follow-up for more than five years has been carried out: the US CREST study.
- The CREST study demonstrated that failures continue to occur even ten or more years after the sterilisation procedure.
- Very few of the studies can be considered population-based: almost all have relied on clinic- or hospital-based recruitment of subjects, and are thus subject to some degree of selection bias (although several have been multi-centre studies spread across several countries, which mitigates this deficiency).
- None of the studies of sterilisation failure have used record linkage methods for follow-up, although three studies have used a population-based record linkage study design to examine the association between tubal sterilisation and subsequent hysterectomy.

CHAPTER 3

Methods: Record Linkage

3.1 Overview of Linkage of Hospital Admission Records

This section describes the methods used to assemble a linked data set from which the incidence of female sterilisation failure could be estimated. Record linkage is used in this study as a relatively cheap (and therefore feasible) means for following up women who are identified in hospital admission records as having undergone tubal sterilisation, on a population-wide basis. Accordingly, this section begins with a brief review of the history, theory and practice of record linkage in epidemiological research. The data sources used in this study and the methods used link them are then described in detail. A critique of those methods and discussion of weaknesses and potential sources of bias in the linked data set is provided.

3.2 The History, Theory and Practice of Record Linkage in Epidemiological Research

There is a large literature on the methodology of probabilistic record linkage. The following subsections are not intended to be an exhaustive review of the subject area, but rather a brief introduction to it, based on the author's personal experience supplemented by the proceedings of an international workshop on record linkage techniques held in 1997.[130]

3.2.1 Definition of record linkage

Record linkage is the process of linking, joining or otherwise associating records (that is, ensembles of data) which relate to the same entity. In epidemiological uses of record linkage the entity is typically a person – usually a patient, but possibly a health care professional, but it can be an institution, or even an address (as in the case of geocoding). Often the records to be linked come from different data collections,

but that is not a necessary condition: record linkage can be undertaken within a single, homogenous data collection.

There are several possible motivations for record linkage:

a) To assemble a richer, or more complete set of data for an individual faster or more cheaply than would otherwise be possible. The additional richness may be cross-sectional, or very often, in the temporal dimension – that is, record linkage is performed to assemble a set of longitudinal records for an individual, which is the case in this study.

b) Geocoding, which is really a special case of a), in which address information is linked to a reference file of addresses which already have geographical co-ordinates or a geographical region (eg a Census Collector District) associated with them, with the aim of determining the geographic location of the original address.

c) De-duplication, where it is known or suspected that multiple records which contain the same, or nearly the same, information about an individual appear in a data set more than once. It is usually undesirable to count such records more than once, hence they need to be removed, or merged into a single record.

3.2.2 Unique entity identifiers or keys

Where a reliable unique patient (or person) identifier (UPI) exists and which covers all of the data sources to be linked, then the process of record linkage is trivial. However, such UPIs are rarer than might be expected. A few countries, such as Norway, issue a unique number for each resident at birth (or on arrival). The number is used to identify the individual in almost all domains, including health care. More common are domain-specific UPIs, such as the UK NHS number, which is (nominally, at least) unique for each client of the NHS throughout England and

Wales (but not Scotland).[131] However, the British social security system uses a different identifier, as do UK police forces, the British tax office and so on.

In NSW, at the time at which the record linkage for this study was done (2001), only two Area Health Services (AHSs), Hunter and Illawarra, had unique MRNs (medical record numbers) for all patients within the scope of the publicly-funded services they provided (although not for all community health services). Empirical investigation of the Inpatient Statistics Collection data used in this study revealed that several other groups of hospitals also used a shared MRN allocation system for the study period – these were:

- Coffs Harbour (H208) and Macksville Hospitals
- Gosford District (B202) and Wyong Hospitals
- Campbelltown (D215) and Camden Hospitals
- Lismore Base (H214), Bangalow and Ballina Hospitals
- Tamworth District (J216) and Gunnedah, Quirindi, Armidale, Glen Innes Hospitals

In all other cases, the scope of an MRN was restricted to the hospital in which it was issued, or to local cluster of associated hospitals (for example, Royal Prince Alfred and King George V Hospitals in Camperdown). Thus, methods of record linkage which do not rely on a unique identifier had to be used.

3.2.3 Beginnings of medical record linkage as an art and a science

Medical record linkage dates back to the publication of two seminal papers by Newcombe and Kennedy nearly half a century ago.[132, 133] In these papers, Newcombe and Kennedy, who were motivated by genetic and biomedical research needs in the atomic energy industry, recognised that in the presence of errors in identifying

information, record linkage becomes a statistical problem – before that, it had been seen as a problem of logical deduction which could be expressed as sets of deterministic rules, such as:

```
If:
     Medicare Number and First name agree;
     OR
     Medicare Number and Surname agree;
     OR
     Date-of-birth and Surname and First
     name initial agree;
     OR
     Month-of-birth, year-of-birth, First
     name, Surname, Wayfare name and Suburb
     agree;
     OR
     Month-of-birth, year-of-birth, Suburb,
     Surname and Initial agree;
     OR
     …etc
Then:
     The records represent the same person.
```

This type of linkage, often referred to as "deterministic record linkage", can be made to work well in many circumstances, although establishing an adequate set of rules can be a very complex can-of-worms, involving a great deal of computer programming.[134]

Newcombe and Kennedy observed that most people would be more likely to consider the second and third pairs of records below to be the same persons, rather than the first pair:

```
John Smith, 10 Herbert St, Dulwich Hill
Jonathon Smith, 12 Alice Rd, Marrickville

Abernatius Rankorata, 23 Glebe St, Glebe
Abenashous Rancaratta, 48 Smith St, Smithfield

Jon Smith, The Oaks, Lightning Ridge
Johannes Smith, Lot 5 Main Street, Lightning Ridge
```

In other words, more weight is given to matching elements when the values of those elements are rare as in the second pair of records above), or where joint probabilities of values are rare (as in the combination of given name, surname and Lightning Ridge, which has a very small population). This led Newcombe and Kennedy to the concept of weights based on the probabilities of chance agreement of partially-discriminating identifying elements (such as name or address components).

The deterministic approach to linkage can be seen as a degenerate case of the Newcombe and Kennedy approach, in which the weights for agreement or disagreement for each data element in each records take fixed values (often assigned heuristically rather than on any statistical basis or from theory). In the simplest cases, as in the example given above, identity weights are, in effect, being used.

3.2.4 The Fellegi-Sunter-Winkler-Jaro linkage model

In the late 1960s, two statisticians working at Statistics, Fellegi and Sunter, formalised Newcombe's and Kennedy's ideas, introducing the notion of "probabilistic" record linkage[135]. In the 1970s, statisticians at the US Census Bureau, most notably Winkler and Jaro, further developed the Fellegi-Sunter model and introduced the use of expectation maximisation (EM) to estimate key parameters in the linkage model.[136, 137, 138]

Details of the theoretical underpinnings are given in Appendix A, but the basic steps involved in probabilistic linkage in the Fellegi-Sunter-Winkler-Jaro model are as follows:

1) Records are assembled in two files to be linked, A and B. In some cases, there is only one file, in which case B=A.
2) Each record in A and B is split up (parsed) into atomic data elements and these data elements are standardised or transformed into a canonical form. For example, if sex is recorded as M or F on file A, and 1 or 2 on file B, then the values on file A are transformed to match those on file B.
3) Each record in file A is paired with each record in file B, and for each pair of records, the individual fields are compared.
4) Comparison of the fields may be on the basis of simple equality, or a string comparator function may be used to provide a measure of the similarity of two names (returning a range of values between 1.0 meaning identical and 0.0 meaning completely different), or a delta function may be used which allows a degree of tolerance between numerical quantities such as age or date. Two ratios are computed based on these comparisons: one for the probability of being a true match given agreement of the values being compared, and one for the probability of being a non-match given agreement (or disagreement) of the values in question. The weights are scaled or weighted in inverse proportion to the frequency of the values being compared i.e. a match on Jelinksy-Jelinsky is given a higher probability than a match on Smith-Smith (in fact, the frequency scaling is somewhat more complicated, as explained below). These probabilities are initially estimated from the data to be linked (in the absence of access to data representative of the underlying population), and then iteratively refined based on the linkage results.
5) The probabilities are summed across all fields for the pair of records, and the ratio of the sums is deemed the “match weight” for that pair of records.

6) Critical values or thresholds for the match weights are determined by expectation maximisation (EM). The theoretical basis for these statistically optimal thresholds or decision rules is covered in more detail in Appendix A. Those pairs of records with a match weight above the upper threshold are deemed matches, those with a match weight below the lower threshold are deemed non-matches, and those with match weights between the thresholds are deemed possible matches (and are usually marked for later "clerical review"). Because the disposition of pairs of records as matched or non-matched effects the matching parameters derived by EM, it is desirable to re-estimate the parameters and re-run the match several times to converge at an optimal set of parameters and frequency scaling values. In practice, this iteration is often done only once, or not at all, particularly when matching very large data sets. It is often not appreciated that the "probabilistic" adjective as applied to the Fellegi-Sunter model of record linkage (and its derivatives) refers to the calculation of this statistically optimal decision rule, and not to the use of match weights to estimate the probability of pairs of records being links or non-links.

3.2.5 Blocking

Unless the two files to be linked are quite small, it is impractical to compare every record in file A with every record in file B. For example, in order to link a file of 100,000 records to a file of 10 million records, one trillion record comparisons would need to be performed.

To overcome this problem, only pairs of records which share sets of certain values are compared. For example, only records in which surname and postcode are the

same are compared. This is called “blocking” because originally it was done by sorting the files and processing them in “blocks”. Modern record linkage software uses indexed retrieval of records to achieve comparison in a sorted order without the need to repeatedly sort the data.

Typically several “blocks” are defined, with phonetically transformed (Soundex) versions of names often used in an attempt to stop spelling and typographical errors from preventing potentially matching records from being compared.

Choice of blocking parameters is important because if too “strict”, records which may be true matches will never be compared, and if too “lax”, an excessive number of comparisons will be performed and the record linkage process will take too long. Theoretically optimal criteria for selecting blocking variables have been described, but are not widely used in practice.[139]

One problem with the use of the Soundex and other phonetic transformations that are commonly used for blocking purposes is that they are very sensitive to spelling mistakes or typographical errors in the initial letter of a name. For this reason, they are sometimes applied to the reversed version of a name and that reversed phonetically transformed version is used in a separate block. For example, two records representing the same person may have surname values of “DORFMAN” and “TORFMAN” in files A and B respectively. The Soundex-encoded versions of these names are D615 and T615 respectively, whereas the Soundex-encoded versions of the reversed names, “NAMFROD” and “NAMFROT” are both N516.

A practical limitation is that commonly available record linkage software permits only a limited number of sets of blocking parameters to be defined (eight in the case of the *AutoMatch*[140] package used for the current study).

Recently several alternatives to traditional blocking have been proposed, drawing on known information retrieval techniques, such as the use of *similarity canopies* or *q-gram indexing*.[141,142]

3.2.6 Parsing and standardisation

With traditional Fellegi-Sunter probabilistic record linkage, best results are obtained if records are broken down into the smallest atomic data elements possible. For example, names should be split up into given names and surnames, and residential addresses should be split up into street number, street name, street type, suburb name, suburb qualifier (north, heights etc), postal code and so on. In this way, it is possible to compare, say, 'Epping North" with "North Epping" by rendering them both as locality=Epping, locality_qualifier=north.

Once disaggregated, values should be transformed to canonical versions: all upper case or all lower case, punctuation removed, abbreviations expanded or standardised and so on. In some cases, it may be useful to cross-compare data elements such as aliases and maiden surnames.

Typically such parsing and standardisation is done using a general purpose programming language (eg SAS data step, BASIC, Perl, Python), or with a specialised tool. For this study, the *AutoStan*[140] software package was used to parse and standardise residential address information (discussed further below). *AutoStan*

allows a series of re-entrant rules to be defined to parse records up into smaller, atomic parts. It has the disadvantage of needing considerable expertise (and time) to programme the rule sets – an excerpt from the *AutoStan* rule set used to standardise the data linked for this work appears in Appendix 2. Only an excerpt can be shown because the complete rule set for street names alone comprises some 8,000 lines of programme code.

Frustration with the time-consuming and difficult task of perfecting the *AutoStan* programmes used to pre-process address data for this study led the author, in conjunction with colleagues at the Australian National University and in the Centre for Epidemiology and Research in the NSW Department of Health, to search for a better method for parsing and pre-processing name and address information for record linkage purposes. A viable alternative approach using hidden Markov models to find the most likely way in which an address or name should be split up was developed in 2002, and a peer-reviewed paper describing the method was published later that year.[143] A copy of the paper appears as an addendum to this thesis, with the author's contribution to the paper and to the work described in it set out on the penultimate page of the paper. The technique developed does not need programming expertise to use, although it does need a set of representative training records from which a model can be learnt. Although it was not used for data preparation in this study, it is now used routinely as part of a geocoding record linkage system within the Centre for Epidemiology and Research.

3.2.7 The statistical theory of probabilistic record linkage versus real-life data

Fellegi and Sunter provided a theoretical basis for determining optimal values for the match weight thresholds but only if conditional independence of the errors in each

field is assumed. In other words, if chance agreement of surname occurs for two records which represent different people, then the probability of chance agreement of the given names for that same pair of records must be completely independent. In practice, this assumption is usually violated to some degree. For example, if two Vietnamese surnames agree purely by chance, then it is far more likely that the given names will also agree by chance (Vietnamese traditionally has only about 35 distinct surnames and relatively few given names). There are methods for estimating the thresholds and other linkage parameters which do not rely on the assumption of independence or errors, but they have not been implemented in any widely available software.

It must also be remembered that the "optimal" matching parameters for given levels of *Type I* and *Type II* error produced by the Fellegi-Sunter model do not necessarily correspond to optimal matching as a human being would see it. This is because a human being tends to focus more on apparent *Type II* errors (false matches) than on *Type I* errors (missed matches). As a result, when record pairs with match weights above the Fellegi-Sunter upper threshold are inspected, there may be pairs which are clearly non-matches, or vice-versa. In these circumstances, many researchers will succumb to the temptation to manually adjust the thresholds. This is perfectly valid if the desire is to reduce the proportion of false links, albeit at the expense of more missed links, and of course the Fellegi-Sunter estimates of *Type I* and *Type II* errors will not be correct.

3.2.8 Clerical review and large files

When large files are linked, the number of pairs of records marked for "clerical review" can be very large (many thousands, or tens of thousands). In the original

Fellegi-Sunter formulation, clerical review was assumed to involve reference to additional information from other sources in an attempt to resolve the question of whether two records refer to the same entity. In practice, such additional information is rarely available (and in certain record linkage architectures, it is not available by design). As a result, the "clerical reviewer" is left to try to divine and implement a consistent set of rules to decide these cases. Newcombe and others have argued that human intuition can be used to extract additional information from these "clerical pairs"[144], while others have argued that this introduces possible bias into the linkage process.[145] Currently available software does not provide any assistance with the clerical review process. A better approach may be to review a sample of the clerical review pairs, devise (or induce) a set of decision rules based on that sample, and then have the computer automatically apply such rules to the remaining "clerical review" pairs.

3.2.9 The transitive linkage problem

One issue to consider when linking records into longitudinal chains is the degree to which the linkage should be transitive. In other words, if record A links probabilistically to record B, and record B links probabilistically to record C, does record A link probabilistically to record C? The answer is "not necessarily". If this is not recognised, it is possible to obtain a chain of records in which, like the children's game of "Chinese Whispers", the identifiers for first record in the chain bear little or no resemblance to the identifiers for final record in the chain.

One solution (and the solution used for this study) is to use the one-to-many matching mode offered by *AutoMatch* and related software. *AutoMatch* uses a linear-sum assignment algorithm to find the best set of matches in the B file for "master"

records in the A file.[138] Exploration of this algorithm is beyond the scope of this thesis. There are several other assignment optimisation algorithms which have been used in other record linkage software, such as the a*uction* protocol in the *Febrl* package.[146]

3.2.10 Assessment of the quality of linkage

Before commencing a record linkage study, it is important to consider what the study demands in terms of the quality of linkage – particularly the acceptable proportions of false links and missed links, and the acceptable biases in those errors. A study which aims to produce an approximate estimate of , say, hospital re-admission rates for a particular disease will have much less stringent requirements in these respects (particularly with respect to bias) than an analytical study looking at the determinants of re-admission. Roos and Wadja provide a practical guide to estimating the feasibility of a linkage study.[147]

Holman, in his course notes on the analysis of linked health data, lists three forms of evaluation of the performance of the record linkage process:[148]

1. Match weight sampling and review techniques, in which pairs of records with composite match weights which place them in the regions before, within and after the "grey zone" of cut-off weights are subjected to detailed clerical review to determine, as far as possible, their true matching status.
2. Chain sampling and review techniques, in which chains of records which have been deemed to relate to the same individual by the record linkage process are subjected to a detailed clerical review to determine logical and biological consistency of all the records in the chain. This is essentially a check for transitive linkage problems mentioned in Section 3.2.9 above.

3. Use of validation data sets, which involves using the same record linkage process on a set of records for which the linkage status of all pairs of records is already known. The record linkage process is, of course, blinded to the "gold standard" linkage status. The known linkage status may be determined through the use of additional information, or it may be because the validation set has been synthesised, as described by Hernandez and Stolfo and implemented in the "generate" utility in the FEBRL record linkage software suite.[149,146]

In the study described in this thesis, methods 1 and 2 have been used and are described in detail in a later section.

Method 3 is also an important method for post-linkage assessment of quality, but is much less commonly performed. Ideally, additional data for a sample of the records being linked is obtained, and this additional data is used to verify both the links and the non-links. The former is straightforward, but verifying the non-links (in other words, checking for missed links) is not so easy, particularly if the intersection between the two datasets being linked is only small. To find all missed links, additional data would be needed on every record in both datasets, not just on the records which have been linked. Thus, if additional data is obtained for only random samples of records from the datasets being linked, then the intersection of the records in the sample will be small, and the number of missed links located will be absolutely tiny, which in turn results in very poor precision and low power for the validation study. This is why the use of synthesised data sets is a useful adjunct, because it allows for an estimation of the missed links rate, at least to the extent that

the characteristics of the synthetic data set reflect the actual study datasets in terms of typographical and other data errors.

McGeechan *et al*.[150] undertook a validation of record linkage between NSW Central Cancer Registry data for women diagnosed with breast cancer in 1992 and NSW hospital admission data for 1991 to 1994, a data set which had been reported on previously.[151,152] The validation study used a 19 percent sample to determine that the linked dataset under-estimated the proportion of women undergoing breast-conserving therapy by about four percent and over-estimated the proportion undergoing mastectomy by one percent. However, no evidence of bias by age or by urban or rural residence was found in the under-estimation of breast conservation. These results are of relevance because the method used to link the data set and one of the data sources are the same as those used in the current study.

Finally, some methods which do not require additional data have been described for estimating the quality of links (but not non-links) in certain circumstances.[153,154] Scheuren and Winkler have also described methods for adjusting for linkage errors in regression analysis.[155,156]

3.2.11 Other approaches to record linkage

The foregoing discussion has concentrated on the statistical model of record linkage which is dominant in the field of epidemiological research. However, other approaches do exist, including several which have come out of information science disciplines.[157,158] Most of the newer methods use "machine learning" techniques, in which the computer learns a model from training data – that is, pairs of records which are considered to be definite matches and definite non-matches (and are

tagged as one or the other). From these, a model is "learned". There is no need to specify the form of the model in the usual sense of statistical modelling – the computer induces both form of the model automatically and estimates the parameters. Two of the most promising techniques involve "maximum entropy models", and "support vector machines".[159,160]

3.2.12 Privacy, confidentiality and architectural considerations

Traditionally record linkage has involved some invasion of privacy, because the person or team undertaking the record linkage process needs to have access to as many personal identifiers, such as name, date of birth and residential address, as possible. Typically this identifying information is associated with medical or other health details. Clearly all possible efforts need to be taken in order to maintain the confidentiality of these data, and to mitigate the loss of privacy by restricting access to the data to as few people as is feasible.

In 1979, two sociologists at the University of Philadelphia, Boruch and Cecil[161], suggested that identifiers be split off from the "substantive" data (such as the medical details) after assigning an arbitrary ID number to each part. The identifiers would then be linked by a team which has no access at all to the substantive data, only to files of names, addresses and dates or birth. The result would be sets of linked, arbitrary ID numbers, which are supplied to the researcher, who then obtains the substantive data-of-interest (but not any of the identifiers, such as name or date of birth), for just those ID numbers, from the original custodians of the data. Variations of this "pseudonymisation" technique were used in Germany in the 1990s for cancer registration[162], and by the Western Australian health data linkage unit, which has published a detailed description of the technique together with suggested

administrative arrangements for the conduct of record linkage projects.[163] Cognisant of the desirability of architectural rather than merely administrative and work practice privacy safeguards in record linkage research, the author has proposed an extension to the Boruch and Cecil model which uses an additional proxy ID mapping entity and public key encryption to hide information from various parties. This proposal, developed during the course of and as a result of candidacy for the degree for which this thesis is submitted, has been published in a peer-reviewed paper.[164] A copy of the paper appears as an addendum to this thesis, with the author's contribution to the paper and to the work described in it noted on its penultimate page.

Some aspects of the Boruch and Cecil model were used in the linkage of records for this thesis: specifically, identifying data were separated from medical details, and linkage procedures were carried out only on the identifying data files. However, one person (the author) had access to both files.

It is possible to undertake record linkage without anyone needing to have access to identifying information (beyond those parties who already have access). These methods, based on one-way cryptographic hashing functions, were described by French researchers in the 1990s and have since been implemented as the main method of compiling administrative hospital admission and other health data in Switzerland.[165,166,167] Improved versions of these techniques, using a variation on the Diffie-Hellman secure key exchange protocol, promise to make record linkage possible without any invasion of privacy.[168,169,170] However, none of these techniques alone provide complete protection of privacy and confidentiality, because the medical and other "non-identifying" data items still need to be assembled for the

purposes of analysis. Unfortunately, an ensemble of non-identifying data items is often quite readily re-identifiable in the presence of additional information and thus cannot be considered anonymous (and thus must still be carefully protected).

3.2.13 Ethical clearance and privacy regime for record linkage studies

At the time that the author performed the record linkage for this study in 2001 and 2002, the use of administrative data sets, for the purposes of research and health services quality assurance, was covered by an internal Privacy Code of Practice for NSW Health published by the NSW Department of Health, as a statutory code of practice pursuant to the *Privacy and Personal Information Protection Act 1998* (NSW).[171] This code specified that written approval by the Director-General of the NSW Department of Health or his or her delegate was required for "internal" linkage, meaning record linkage within or between one or more data collections owned and operated entirely by the NSW Department of Health, and that clearance by an ethics committee was not required. This study internally links a single NSW Department of Health data collection. Written permission to undertake the study and perform the internal linkage was obtained from the Chief Health Officer and Deputy Director-General, Population Health, who was delegated to provide such permission. Approval was given subject to the conditions that no identified or re-identifiable data leave the NSW Department of Health premises at North Sydney – hence all record linkage work had to be performed on-site – and that no individuals or institutions would be identified in any published or unpublished theses, papers or reports based on the linked data set. These conditions have been strictly observed in the preparation of this thesis.

Subsequently, the *Health Records and Information Privacy Act 1992 (NSW)*, and statutory guidelines pursuant to it, which came into force in September 2004, require that record linkage studies obtain clearance from a properly constituted ethics committee before they proceed.[172,173] These requirements are not, of course, retrospective. The rationale for the new requirement is that, when performed in the traditional manner (that is, not using the privacy-preserving technologies described in the preceding section), the privacy of individuals is necessarily invaded to a small degree by investigator(s) who have the ability to see both the identity of the study subjects and/or their medical details. Usually these medical details are from more than one source, and therefore the investigators may have access to an ensemble of information which hitherto may have only been known to the study subject but not to any one of their medical attendants or health care providers. For these reasons it is vital that investigators minimise this potential or theoretical invasion of privacy through strategies such as avoidance of the simultaneous examination of identities and medical details. Ethics committee consideration has therefore deemed to be required to determine whether the unavoidable, although slight, invasion of privacy involved such record linkage studies is outweighed by the "public good" which will accrue from the research. Under the former arrangements in NSW, the Director-General of the NSW Department of Health effectively made this determination on behalf of study subjects, at least for studies in which only Department of health records were used (as described here). It should be noted that it is rarely feasible to obtain individual consent in record linkage studies, due to the number of subjects involved.

In order to maximise the protection of privacy, the following additional measures were observed for this study:

- All record linkage work was undertaken solely by the author, and no other person was given access to the files involved
- As mentioned in the section above, the principles established by Boruch and Cecil[161], and later promulgated by Kelman *et al.*[163], in which the directly-identifying and partially-identifying data items (described in a subsequent section) used to carry out the record linkage were split off (into a separate file) from the substantive medical details for each record, were observed. An arbitrary unique identification number was used to link the identifying details to the medical details and files were stored separately and in encrypted form. Thus, the phase of record linkage which required access to these identifiers was performed using files which carried no information about the medical details of each person. After this phase was complete, the medical details were then linked using "pointers" derived from the first phase (more details are given below).
- When not being actively worked upon, all files containing identifying and partially-identifying data were strong encrypted using GPG encryption software and a cryptographically strong key.[174]

3.3 Data source

The data used in this study were drawn exclusively from the NSW Inpatient Statistics Data Collection (ISC).[175] Since July, 1993 this data collection has been a complete census of all admitted patient separations from New South Wales public acute care hospitals, public psychiatric hospitals, public multi-purpose clinics, private hospitals, and private day procedures centres. It does not include nursing homes. Separations, which imply a formal hospital admission, include discharges, transfers, extended leave (typically from long-stay psychiatric institutions) and deaths in hospital.

Patients seen only in emergency departments, outpatient clinics, community health centres or other "ambulatory" settings are not included, unless they are also admitted to a hospital bed. However, patients undergoing most forms of surgical procedure, even on a day-only basis, are included in the data collection. All surgical procedures requiring a general anaesthetic are captured. Data are collected on a financial year (1st July to 30th June) basis.

Prior to July 1993, data from some smaller private hospitals were collected on a sampled basis - typically two months of data collection, followed by two months of no data collection, and so on. However, 97% of hospital separations in NSW were enumerated in the 1992/93 financial year ISC. This minor degree of sampling in that year of data has been ignored in this study and is not expected to have a significant effect on the results.

Although records for NSW residents who were admitted to interstate hospitals were available, they were not used in this study because the partially-identifying demographic data items used to link records, such as address and date of birth, were not available for these records (the data items in question are not supplied to the NSW Department of Health by health authorities in other States and Territories, and vice-versa).

The ISC data files used for this study were accessed via the HOIST population health data warehouse system, which was designed, established and built by the author in the 1990s in the predecessor to the Centre for Epidemiology and Research in the NSW Department of Health, and which continues to be operated by the Centre.[176]

ISC records consist of demographic data items, administrative data items and coded information on diagnoses related to and procedures performed during a particular admission to hospital. The availability of identifying and/or partially-identifying data items is of particular importance for record linkage studies. The ISC contains a number of data items which have various degrees of discriminating power for personal identity. These include:

- Hospital identifier code plus patient Medical Record Number, which, nominally at least, uniquely identifies an individual within the context of a single hospital, or, for two NSW Area Health Services (Hunter and Illawarra) for the latter part of the study period, within all public-sector hospitals in that Area Health Service.
- Sex
- Date of birth
- Full residential address, including street number and name, suburb or town and postcode
- Country of birth
- Language spoken at home
- Marital status
- Aboriginality
- Insurance status

A description of the manner in which these data items were used for record linkage in this study appears below. Importantly, name information was not available (in 2003 it became incompletely available for years since 2001 but ethical and administrative arrangements for it use in record linkage studies were not resolved within the NSW Department of Health until 2005 and thus name information could not be used for this study). Of course, a significant proportion of women change their

surname after marriage, which would reduce the reliability of such linkage, but not entire frustrate it, because the given names and other partial identifiers remain unchanged. In addition, the majority of children are born after marriage (and any associated change of surname), although obviously not all. However, even in circumstances in which women change their name after marriage (or change it back after divorce or separation), the availability of a given name would still add considerable power to the linkage process. Additionally, linkage with marriage certificates, which are public documents and are available in electronic form might further reduce the impact of marriage-related surname changes –an array of one or more surnames could then be assembled for each woman and the Cartesian product of this array used in the surname comparison. The linkage software used for this study has facilities to handle such arrays of alternative names or aliases.

3.3.1 Commentary on data sources

Death is an obvious censoring event in any follow-up study (and is the outcome of interest in survival studies, of course). Although the ISC data contains information on admissions terminating in death while in hospital, deaths occurring outside a NSW hospital are not included in the ISC and thus will not be picked up in a linked data set based solely on it. One way of avoiding this problem is to additionally link death certificate or death index data. This additional step was not done as part of this study because of the additional time and resources involved, but would typically be done as part of research for a higher degree. However, as noted in Section 2.7 above, the all-causes death rate in women of reproductive age in Australia is quite low (under 70 deaths per 100,000 women per annum), and the majority of deaths which do occur in this age group occur during an admission to hospital, even for death due to trauma after, say, a motor vehicle accident.[117] Thus the failure to include death

certificate data in the linked data set is not expected to have had a major impact on the results obtained, particularly given that death is, in the context of this study, used only as (relatively rare) censoring event, and not as an outcome event. There are, unavoidably, far more important gaps in the completeness of follow-up in this study – these are discussed in greater detail in the next section and subsequently.

3.4 Record linkage study design

The primary aim of the epidemiological study described here was to estimate a lower bound for the cumulative incidence of failure of surgical (female) tubal sterilisation procedures using probabilistically linked records of admissions to New South Wales hospitals. The secondary aim was to explore the effect of various factors on the estimated failure incidence using survival analysis techniques.

The design is a retrospective passive follow-up study of a dynamic population, also known as an "open cohort". The study population is dynamic, rather than a closed cohort, because subjects who enter the study population may subsequently leave, or be excluded from it because they are no longer at risk of the outcome events of interest. Many studies with designs similar to this one are commonly referred to simply as retrospective cohort studies, and the term "cohort" will be used subsequently in this thesis, for the sake of brevity. Follow-up is "passive" because no subjects were contacted as part of the study, and their subsequent status following entry into the study population was determined entirely by record linkage of hospital admission records relating to them.

For the purpose of this study, two sets of hospital admission records were drawn from the NSW ISC records for the financial years 1992/1993 to 1999/2000 inclusive.

The first set comprises records for women who underwent any form of surgical sterilisation procedure during a hospital admission – the details of exactly which records were selected appear below. This set of records was used to form a cohort of women who have undergone surgical sterilisation in a NSW hospital. It is not, *per se*, a cohort because the records relate to hospital admissions, (or more precisely, hospital separations), not persons, and one person may have multiple admissions for a sterilisation procedure, although this is rare. However, a side effect of the record linkage process is that multiple sterilisation procedure records for a given person can be effectively "condensed" or "collapsed" into a single record for that person, thus forming a true cohort (or, more strictly, a true dynamic population).

The study outcomes of interest are any form of conception subsequent to a surgical sterilisation procedure. However, it is also necessary to detect censoring events occurring subsequent to the first surgical sterilisation procedure that a woman undergoes. Censoring events must include anything that indicates that she is no longer at risk of the outcome of interest and should therefore no longer be considered to be under observation or subject to (passively, in this study) follow-up for outcomes of interest. Censoring events include any procedures or events, including presumed menopause based on calculated age, and death, which indicate that a member of the study population is no longer at risk of conception (previous sterilisation procedures notwithstanding, because we are testing the efficacy of that method of contraception), or any event or procedure which indicates an attempt to reverse or bypass the original sterilisation procedure. By extension, any conception outcomes occurring after a censoring event must be disregarded for the purposes of analysis.

From these data, through the process of record linkage described below, information on the total person-time-at-risk of subsequent conception in women who have undergone surgical sterilisation can be assembled, together with information on the conception outcomes of interest occurring in these women.

The foregoing is, of course, an idealised conception of this study, and in reality there were several deficiencies in the linked data. The first was that the record linkage process itself was necessarily imperfect. This is partly due to methodological flaws and incorrect theoretical assumptions, as discussed in Section 3.2.7 above, and which are common to all studies which use the dominant Felligi-Sunter probabilistic record linkage paradigm, and partly due to data errors. Secondly, there were gaps in the partially-identifying data items available – in particular, the absence of any name information. Thus, some records which ought to have been linked will not have been— resulting in missed outcomes and missed censoring events—and some records will have been linked incorrectly. The third, and perhaps more serious, problem is that the follow-up data used to detect outcomes of interest and censoring events was not complete, neither in time nor in geographical extent. Incompleteness in time was accommodated at the analysis stage by censoring at the last date of data availability. However, incompleteness in the geographical coverage of the available data (NSW hospitals only) means that women who move interstate, live temporarily overseas or emigrate permanently were "lost to follow-up". Loss to follow-up is an important issue in any longitudinal study looking at measures of survival or failure (just fact-of-failure and also time-to-failure). However, in studies using active follow-up methods, it is usually possible to determine, at least approximately, when study subjects are no longer under observation – typically attempts at mail or telephone contact with subjects fail, or they cease to attend a particular clinic or

hospital if the study is based in one or more health care facilities. These subjects can then be censored at the time of their last contact. However, in passive follow-up studies, such as those that use record linkage, it is not known if particular subjects are no longer "under observation" – that is, whether subsequent outcome or censoring events for those subjects are not able to be detected because the subject may have relocated outside of the geographical scope of the record linkage data. The result of this problem of unknown extent of loss-to-follow-up is that the numerator outcome events tend to be under-ascertained, and the denominator person-time at-risk-and-under-observation tends to be overestimated. The result is that both unconditional and conditional cumulative incidence of the events under study (which may be failure events, as here) tends to be underestimated. As noted previously, this is the reason that the current study could estimate only the lower bound for cumulative incidence of failure. However, this inherent tendency towards underestimation of cumulative incidence of failure in studies involving follow-up by record linkage, or other means of passive follow-up, does not necessarily extend to the estimation of ratios or other comparative or relative epidemiological measures from such studies. Under-ascertainment of outcome and censoring events and overestimation of time-under-observation are measurement errors which may be differentially distributed, and thus relative risks, hazard ratios or odds ratios for comparisons between groups may be biased in either direction – either towards or way from the null, despite the fact that absolute measures of incidence will tend to be biased towards zero.

The size of the "unknown loss-to-follow-up" problem can be estimated from estimates of the missed link rate and from estimates of out-migration from the scope of the record linkage follow-up data. In the Australian context, out-migration from Western Australia, where there is a well-established population-based record linkage

facility, might be expected to somewhat lower than for NSW, but will not be zero.[177] Such estimation is rarely done, and is beyond the scope of this work, but the issue is duly recognised here. Interestingly, as mentioned in Section 2.5.3 above, Lassner *et al*. examined this issue by using active follow-up on a 15 per cent random sample of their study population (although they were not using record linkage for the purposes of passive follow-up).[64] They found no difference in ascertainment of outcomes (pregnancies) between the active and passive follow-up groups. It is unclear how generalisable this finding is to other studies and settings. Cohort studies using mixed active and passive (through record linkage) follow-up methods, such as the NSW "45 and Up" study, may have a theoretical advantage when it comes to the estimation of absolute epidemiological measures such as incidence, because they may be able to use the failure to contact subjects through active follow-up as an indication that those subjects are also effectively lost to passive follow-up through record linkage, and thus use this information to censor the contribution to the denominator person-time at risk made by those subjects.[178]

As already mentioned in Section 3.3.1 above, the ascertainment of outcome and censoring events could also be improved by linkage with other data collections, notably with: a) death certificates, which would provide data on deaths occurring outside NSW hospitals; b) births data from the NSW Midwives Data Collection, which would provide data on home births as well as confirmation of births occurring in hospitals; and c) the NSW Emergency Department Data Collection, which might provide additional data on miscarriages managed on an outpatient basis. In addition, several assisted conception data collections which have been started in NSW could also be used to provide extra censoring data on assisted conception attempts. Linkage with these additional data sources was not undertaken as part of this study because of

the very considerable additional time and effort that such a multi-way linkage would have taken (remembering also that this study, unlike many other record linkage studies, involves both the linkage of the source data and its analysis by a single individual). However, subsequent to he linkage work for this study, the NSW Centre for Health Record Linkage (CHeReL) was established.[186] All of the foregoing data collections, with the exception of assisted conception registers, are now routinely linked into the CHeReL "Master Linkage Key". This means that, should the current study be repeated, it will be very much easier to also use these additional sources data on outcome and censoring events.

It should also be noted that since the mid 1990s, there has been increasing interest in and evaluation of the medical management of first trimester spontaneous abortion on an ambulatory (non-admitted) basis using agents such as misoprostol (for example, the trial reported by Shankar *et al*.).[179] Determination of whether such clinical trials had a significant effect on the clinical management of abortion in the later 1990s is beyond the scope of this study, because it would involve linkage between emergency department and admitted patient data, as well as an examination of the pharmaco-epidemiology of misoprostol use. Nevertheless, it is an addition potential cause of under-ascertainment of outcome events, in the absence of linkage with ambulatory care data.

3.5 Overview of data preparation and record linkage procedures

The data preparation and record linkage phases of this study can be summarised as follows. Further detail on some of the steps is provided in subsequent sections.

1. Permission to access and use the data source, which is a secondary collection of administrative records for almost all public- and private-sector hospital admissions in New South Wales, for an eight year period from July 1992 until June 2000, was obtained as described in Section 3.2.13 above.

2. A set of hospital admission records, each of which contain one or more ICD-9-CM or ICD-10-AM codes indicating that a tubal sterilisation procedure was performed during that admission, was selected. Details appear in Section 3.6 below. Any of these records which were not for females between the ages of 12 and 59 years were discarded. This age range might be considered an overly-broad definition of reproductive age, but it was selected to allow for up to ten years of follow-up in the eldest study subjects, and is subject to further age filtering at the analysis stage. This set of records, denoted (by record linkage convention) as File B, provided the basis for the cohort to be followed up, although at this stage there was more than one record for some subjects (because the file was of hospital separations, not of persons).

3. A second set of hospital admission records, each of which contained one or more ICD-9-CM or ICD-10-AM codes indicative or suggestive of: a) sterilisation failure (that is, conception); or b) a censoring event, was selected. Any of these records which were not for females between the ages of 12 and 69 years were also discarded. Censoring events comprised: i) procedures which imply permanent sterility even if the Fallopian tubes are patent and presumably functioning; ii) attempts at reversal of tubal sterilisation; iii) attempts at assisted reproduction which bypass or which do not rely on functioning Fallopian tubes; or iv) death due to any cause during a hospital admission. Details of the codes

used to select these records are given in the next section. This set of failure and censoring event records, known as File A, provided the means to undertake passive population-based follow-up, through record linkage, of the pseudo-cohort in File B.

4. Records in File A (failure and censoring events) were linked, using a range of personal partial identifiers (such as date of birth and address), to records in File B, in an exclusive many-to-one fashion. This allowed each member of the pseudo-cohort in File B to be linked to multiple failure and censoring records in File A. However, each failure or censoring event in File A was permitted to link with only one member of the cohort in File B.

5. As part of the process of linking File A to File B, records in File B were also linked to one another. This allowed, at least in theory, all the sterilisation records relating to a single person to be grouped together, thus allowing the pseudo-cohort in File B into be later converted into a true cohort of person-based records. As expected, it was rare for a person to undergo more than one sterilisation procedure – details are given in Chapter 4.

6. All the records from both File A and File B which were linked together were re-assembled into a person-oriented set of hospital admission events. Such sets will be referred to here as "match sets". By definition, each match set contained at least one sterilisation event, and zero or more failure and/or censoring events. The events within each match set were sorted into chronological order, and all failure and censoring events which occurred prior to the first sterilisation event in that match set (that is, for that person) were discarded.

7. Additional censoring events were added to each match set for the date on which the subject reached 47 years of age (an arbitrary end-of-fertility censoring point) and for 30th June 2000 (the end-of-follow-up censoring event). Therefore every match set contained at least one censoring event.

8. A final "analysis" dataset was extracted which contained, for each person in the study population (the "cohort") the nature and dates of: a) the sterilisation event; b) the earliest failure (if any) or censoring event to have occurred. The age and health insurance status of the subject and the identity of the hospital in which the sterilisation procedure was performed were also included in this dataset.

9. Descriptive, life-table and other survival analysis was then performed on this dataset (described in detail in Chapters 4 and 5).

Unless otherwise stated, all data manipulation and some of the analysis was performed using version 8.2 of the SAS System software, and all record linkage processing was undertaken using *AutoStan* and *AutoMatch* software from Matchware Technologies.[180,140]

3.6 Selection of records for linkage files A and B

For the financial years 1992/93 to 1997/98, admission-related diagnosis and procedure information was recorded in the ISC using the first, second and third Australian editions of ICD-9-CM (International Classification of Diseases, Version 9, Clinical Modification).[181] ICD-10-AM (International Classification of Diseases,

Version 10, Australian Modification) was used from 1998/1999 onwards.[182] The codes used to select records are given in Appendix C.

ISC records for the 1992/1993 financial year contained only five fields for diagnosis codes for each hospital separation, and only four procedure code fields. Records for subsequent years contained 11 diagnosis code fields and ten procedure code fields. Not every available code field was filled with data in every record – indeed, most records contained only one or two diagnosis codes and zero or one procedure code, although some records contain many more codes. Nevertheless, in each year, all available diagnosis and procedure codes fields were searched for the target codes appropriate for that year of the ISC data collection (as listed in Appendix C). There is a potential source of bias in the time-dependent transition from fewer to a greater number of diagnosis and procedure code "slots" in administrative data collections.[183] However, it is unclear whether restriction of the number of codes considered for later years of a data collection (a "lowest common denominator" approach) helps to correct this potential source of bias. Such restriction was not used in this study.

If a record contained one or more of the target diagnosis or procedure codes, it was included in the selected subset, only once – if a record contained two or more target codes it was not, of course, selected twice. Records were additionally filtered for sex (females only) and age 12 to 59 years at time of separation.

The total number of records in each year of the ISC and the number of records selected for files A and B are shown in the following table.

Table 6 - Numbers of ISC records for A and B files

Year of ISC data	Total number of records searched	Number of records selected for File A	Number of records selected for File B

1992/1993	1,525,040	208,379	9,586
1993/1994	1,703,557	223,712	9,794
1994/1995	1,745,566	219,539	9,162
1995/1996	1,824,688	217,863	9,203
1996/1997	1,859,510	217,429	8,098
1997/1998	1,923,146	215,520	7,809
1998/1999	1,902,826	207,738	7,314
1999/2000	1,909,649	205,075	6,528
All years used in this study	14,393,982	1,715,255	67,494

Following additional consultation with experienced medical record coders (Ms Susan Travis and Ms Christine Ewart in the Centre for Epidemiology and Research, NSW Department of Health), the author decided that some of the codes were insufficiently specific to be used to infer a conception outcome. Therefore, in accord with the goal of this study of deriving a conservative, lower-bound estimate of the incidence of failure, these codes, which are annotated in the table in Appendix C, were not treated as denoting a failure in the final data set used for analysis.

The SAS programme code needed to effect these selections was necessarily rather lengthy and complex, due to the large number of diagnosis and procedure codes being searched for in multiple fields in each record, and was thus potentially susceptible to subtle errors. The programme code was reviewed repeatedly for possible errors, but in order to provide an independent quality assurance check, the following additional steps were taken:

1. An independent programme was written which assigned each of the target codes in the data sets for files A and B to one of four “endpoint” categories:

a. 1 = conception (failure);

b. 2 = no longer at risk of pregnancy (apart from the original tubal sterilisation procedure) (censoring);

c. 3 = reversal of sterilisation (censoring);

d. 4 = assisted reproduction bypassing the Fallopian tubes (censoring).

2. For each target code in the selected records, a record was written out to a temporary quality assurance dataset, which therefore then contained one record per target code, not one record per hospital admission (possibly each with more than one target ICD code).
3. The individual diagnosis and procedure codes in the quality assurance dataset were then cross-tabulated against the endpoint categories, and a manual check was made to ensure that no unexpected codes had been selected, and that instances of every expected code were present. This effectively validated the initial record selection statements (the SAS "where" clauses), as well as validating the endpoint assignment programme statements.
4. The validated endpoint assignment statements were then re-used in step 8 as described in Section 3.5 above.

In the course of looking up the ICD-9-CM codes for tubal sterilisation procedures, the author discovered that the Australian ICD-9-CM manuals list the incorrect code for laparoscopic tubal sterilisation in their index volumes. The correct codes are given in the tabular ICD-9-CM volumes. Discussion with experience medical coders (Ms Susan Travis and Ms Christine Ewart) revealed that most coders, for reasons of speed, used the index volume in preference to the tabular volume, and thus it is quite likely that a significant proportion of laparoscopic tubal sterilisations were inadvertently coded as open procedures.

There is, of course, the fundamental issue of how well diagnoses, procedures and other events which patients experience while in hospital are reflected in the codes recorded in the source data collection. Several studies addressing this issue have been published. Henderson *et al*. compared hospital admission records coded especially for auditing purpose with the routine coding of the same records in Victorian hospitals in the late 1990s. They found no appreciable differences in the accuracy or extent of coding of diagnoses or procedures.[184] However, MacIntyre *et al*., in a similar study, also in Victorian hospitals, in 1993 and 1994, found significant differences between audit coding and routine coding of a random sample of hospital admission records.[185] There are no extant studies for NSW hospitals. A comprehensive review of the literature regarding the accuracy and completeness or the routine, administrative coding of hospital admission records is beyond the scope of this study, but it must be noted that an intrinsic limitation of this study, and almost all other studies which rely on hospital administrative data collections, is that the quality of the data coding is either unevaluated or, at best, only weakly characterised. From an epidemiological perspective, under-ascertainment of outcome and censoring events due to inadequate coding of the administrative data could be expected to have a similar effect on the study results as loss to follow-up due to the limitations of the record linkage process, as discussed in Section 3.4 above.

3.7 Record linkage using *AutoMatch*

Plain text ("ASCII") files containing only the identifying data items needed for record linkage and a unique record identification number were written out. Some pre-processing of the textual address data in these files was done to remove extraneous characters and to concatenate the address fields into a single string. Although the ISC

nominally contains separate fields for street address, suburb and postcode, address information often overruns these fields and ends up being stored in one of the other address fields, as a result of limitations and programming errors in the source hospital information systems. Experience has shown that it is better to therefore re-join all these fields and then parse them apart using special-purpose address parsing software. The ASCII files were then processed with *AutoStan*, as described in Section 3.2.6 above. *AutoStan* splits the address information into "atomic" elements, such as street number, street name, street type and so on, and also "standardises" values into canonical forms – for example, "st", "str" and "strt" are all converted to "street". In addition, Soundex (phonetically encoded) representations of street and suburb names were added to the files by the *AutoStan* scripting programme (written previously by the author and colleagues) for the purposes of "blocking" in the record linkage phase. Because Soundex codes are not robust to errors in the initial letter, Soundex codes of the reversed street name and suburb names were also added. This has been found to be a useful heuristic in past record linkage exercises.

After the files A and B were pre-processed in this manner, they were probabilistically linked using *AutoMatch* software. A complete description of all the steps involved would be somewhat nugatory and thus only the salient steps will be described below.

AutoMatch requires that a linkage specification file be written which sets out how records will be "blocked", that is which sets of records will be brought together, pair-wise, for more detailed comparison, and then for each block, which fields will be compared and the nature of those comparisons. An annotated and abridged copy of the *AutoMatch* linkage specifications file created for this study is provided in

Appendix D. Eight sets of "blocks" or matching passes through the data were employed (the maximum number allowed by the software), with records sharing the same MRN within the same hospital compared and matched on that criterion in the first block – essentially a deterministic match on hospital code plus MRN. The second pass compared records with the same MRN within a "pseudo-hospital code" created so that all hospitals in the two Area Health Services (AHS) (and the handful of other hospitals listed in Section 3.2.2 above) with shared MRN allocation systems had the same code, and thus records with same MRN for any of the hospitals in those two AHSs or hospital groups match, also deterministically. Subsequent passes used more traditional probabilistic record linkage blocking strategies, comparing on various combinations of date of birth, day, month and year or birth, postcode and Soundex and reverse Soundex phonetic encodings of suburb name and street name. The intention was to throw a wide "net" on each pass, with the nets overlapping, so that the likelihood that a pair or records which might possibly match never being compared to one another in one or other of the passes is exceedingly small.

Apart from the first two passes, in which record pair comparisons were essentially deterministic, based on hospital code and MRN, records were compared on the basis of date of birth, country of birth, and residential address information. Marital status was not used in this study because the overwhelming majority of study subjects had marital status recorded as "married" – perhaps unsurprisingly, as tubal sterilisation is not a popular method of contraception amongst younger and/or unmarried women.

After each pass, pairs of records with a range of composite "match weights" were sampled and compared until approximate upper and lower cut-off values were determined – the values below, between and above which pairs of compared records

were deemed to be non-links, possible links and links. After setting these heuristically-determined cut-off weights, the linkage pass was re-run and "clerical review" on all pairs of records which were possible links (in the "grey zone") was performed. Typically several thousand pairs of records needed to be reviewed in this fashion, a process which took an average of about 10 hours per pass (bench notes indicate that the author spent a total of 76 very tedious hours doing clerical review for this study). A set of consistent decision rules was developed for the purposes of deciding whether pairs of records were likely to be the same person. In practice, in the absence of additional information from external sources, such human-mediated clerical review may not add a great deal to the overall quality of the linkage, and may introduce biases when different clerical reviewers apply different sets of decision rules (which are often not codified), or the same reviewer applies the rules inconsistently. Nevertheless, clerical review remains part of standard probabilistic record linkage practice.

After all eight passes were run, the *AutoMatch* MPROB programme was used to update the m-probabilities, using an expectation maximisation algorithm, and each pass was then repeated. Fortunately *AutoMatch* includes a facility which allows previous clerical review decisions to be automatically re-applied to the same pairs of records. Therefore only a small number of additional clerical reviews were needed after each of these second set of linkage passes through the data using the optimised match probabilities for each value in each field.

Finally, the *AutoMatch* data extraction utility was used to write out a single file containing all records from both the A and B files, with linked records identified by a common "match set" number.

3.8 Post-processing of linked data

At this point, a file containing pointers to records for hospital admissions belonging to, nominally at least, the same person had been created. The set of records pointed to for each person contained at least one admission for a surgical tubal sterilisation procedure, and zero or more admission records for other tubal sterilisation procedures (from file B) and/or zero or more admission records indicative conceptions or censoring events. However, none of these events were ordered in time, and the conception-related or censoring events may have preceded the surgical sterilisation event or events – the record linkage was carried out without regard to temporality, because *AutoMatch* has no facilities for temporal sequence comparisons, only temporal difference comparisons. Therefore a series of post-processing steps were required:

1. Each record was flagged to indicate whether it contained a sterilisation procedure event, an event indicative of a conception, or a censoring event (including in-hospital death and presumptive menopause as noted previously), using the previously validated SAS programme statements discussed in Section 3.6 above. It was possible for each record to have multiple flags – for example, a record might contain a sterilisation procedure flag and a death-in-hospital flag should the women have died during the procedure (this circumstance did not, in fact, occur in the study data set). As noted in Section 3.6 above and in Appendix C, some of the procedure codes that were initially selected were, after review, felt to be insufficiently specific and were therefore not used to infer conception in the final data set used for analysis.

2. Within each match set (that is, the set of records belonging to a putative person), the hospital admission/separation records were sequenced in order of date of separation, and the index (that is, the first) admission record for a tubal sterilisation procedure was identified.
3. All admission records which pre-dated the index tubal sterilisation procedure in each match set were discarded.
4. All match sets containing an outcome event, that is, a diagnosis or procedure code indicative of a conception, were printed out and the quality of the linkage between the index sterilisation record and the outcome record was manually reviewed. All outcome records which appeared to have been erroneously linked to the index sterilisation were discarded, using the same decision rules used for previous clerical review. This was done as an additional check to minimise the number of false outcome links in the data set. This was necessary because, as discussed earlier, the putative biases in this study due to the limitations of the data and linkage process were towards under-ascertainment, and thus it was deemed wise to minimise any possible bias in the opposite direction which might occur due to false links. Approximately 1,000 match sets were manually reviewed, a process which took approximately 50 hours, with 43 outcome events eliminated as a result. Ideally the same review process would be applied to censoring events within match sets, but because there were over 65,000 such sets, such manual review was infeasible. It is theoretically possible that false links with censoring events could incorrectly reduce the person-time in the failure rate denominator, and thus bias the resulting estimate of the incidence of failure in an upwards direction. However, as discussed above, and further discussed below, it was felt that data and linkage limitations would make missed links far more

common than false links, thus ensuring that information bias would be towards zero.

5. All records indicating a hospital admission from which conception could be inferred and having a date of admission within 28 days of the index sterilisation procedure were assumed to result from luteal phase failures or from failure to test for pre-existing pregnancy prior to the procedure. Such "pruning" of outcome events which may have been due to a luteal phase pregnancy or may have been due to a very early failure (in the first ovulatory cycle after a sterilisation procedure) was, again, "erring on the side of caution", in an effort not to bias the estimate of failure incidence upwards. One hundred and twenty-three outcome events were removed as a result of this conservative "pruning". It must be noted that this pruning may not have removed conception results for all women who were already pregnant prior to their sterilisation procedure – it merely ignores failure outcomes in women who may have only just conceived at the time of the sterilisation procedure and who may not yet have missed a menstrual period or experienced any other signs or symptoms of pregnancy, or who might not have returned a positive beta-hCG test result when screened for pregnancy prior to their tubal sterilisation surgery.
6. All admission records subsequent to the first censoring event within each match set were discarded.
7. For match sets which did not contain a censoring event, a censoring event representing end-of-observation (end-of-follow-up) was added to the last record in the match set, with the censoring date set to 30th June 2000, which was the last date for which data was available.
8. The final dataset was "rolled-up" (summarised) by creating a data set with a single record per putative study subject. This summary data set contained details

of the index sterilisation event, the first occurring censoring event (including death or end-of-observation on 30th June 2000) and the first occurring outcome event. The date of outcome was deemed to be the date of admission to hospital for the record containing the outcome event. Ideally, the date of conception would be used in any time-to-failure analysis, but this information was not available. For outcome events such as term deliveries, the date of conception could be estimated, but for other outcome events such as spontaneous or induced abortions, duration of gestation information was not available nor could it be easily estimated. Such estimation of date of conception is a refinement which might be undertaken in subsequent studies of this type.

3.9 Comments on the record linkage process used in this study

Apart from MRN within the same hospital (or Area Health Service), the main data items on which linkage was based were residential address and date of birth. As mentioned above, name information was not available on ISC data records at the time the linkage for this study was performed, although it has subsequently become available and is now used as part of the ongoing probabilistic record linkage activities at the recently established NSW Centre for Health Record Linkage (CHeReL).[186] Unfortunately date of birth, although highly discriminating, is not quite sufficiently discriminating in the context of hospital admissions to be useful for matching on its own. Therefore, despite the probabilistic record linkage methods used, the matching in this study essentially reduces to matching on MRN within a hospital, or matching on a combination date of birth and residential address. The latter is quite specific and would be expected to result in very few false links – even within a large block of apartments, the probability of two people sharing a date of birth is less than 0.01 (assuming even probability of approximately 35,000 possible

dates of birth and up to 350 people living in such a dwelling). This probability of false linkage is further reduced in the context of this study by the additional requirement of having been admitted to hospital for any of the target procedures or conditions. For these reasons, it is expected that false links will be quite rare in final study data set, certainly less frequent than 1 in 100 and possibly one or two orders of magnitude less frequent than that.

However, missed links are rather more likely. Bell used life tables derived from Australian Bureau of Statistics household survey data from 1987 to 1992 to estimate how often Australians change their residential address.[187] He found, unsurprisingly, that younger and/or unmarried people move more often, and that middle-aged people and those living as part of a family unit move less often. He estimated the median time between moves for 30 to 44 year olds was 4.3 years, and for couples, 6.6 years. However, if we assume that 30 per cent of the study cohort move within a five year period (which is consistent with Bell's estimates, given the highly skewed, log-normal frequency distribution of changes in residential address that he found), or about four per cent each year (with independent probability of moving from one year to the next), then the probability that a subject retained the same address over an eight year period is 0.96^8 which is 0.71. Therefore as many as 30% of links may have been missed in this study due to changes in residential address. Although this estimate may be unduly pessimistic, when other causes of missed links, such as data recording errors (for example, incorrect dates of birth) and data scope issues, such as admission to interstate hospitals, are considered, it is clear that a substantial proportion of links may have unavoidably gone undetected. Of course, these missed links will result in missed censoring events as well as missed failure events, but in either case the effect, as noted previously, will be to depress the estimated incidence

of failure and thus the estimates produced by this study must be considered lower bounds for the true values.

CHAPTER 4

Results

This Chapter gives a brief description of the final data set used for analysis and the variables included in it, before proceeding to present descriptive univariate analyses, selected bivariate results, and estimates of the cumulative incidence of failure. Tables in this chapter were generated using SAS, although the free, open source R statistical package was used to generate all graphics due to its superior capabilities in this respect.[188]

4.1 Numbers of linked records

Thirty-seven thousand and seventy-eight of the 67,494 sterilisation admission records linked to 82,787 of the 1,715,255 endpoint admission records. Records were grouped into person-oriented "match sets" and pruning as described in Section 3.8 above was carried out, resulting in a final data set containing 65,499 such match sets. This was then summarised into a final analysis data set also containing 65,499 records. Some 745 women (approximately 1.1 per cent) were recorded as undergoing a second sterilisation procedure – these second procedures were ignored in the analysis and will not be examined further in this thesis, although could be further investigated in a more exhaustive study. Sixty-eight per cent of the women undergoing a second sterilisation procedure also experienced sterilisation failure (that is, they appeared to have conceived subsequent to the first sterilisation procedure).

Although there are a large number of variables available in the ISC, only a small number which were considered to be most highly relevant were retained in the final analysis data set in order to keep the scope of the analysis task manageable. These variables were age and age group at the time of the first sterilisation procedure, the type of sterilisation procedure, the code for the hospital in which it was carried out, whether the woman was admitted as a private or a public patient at the time of the

procedure, her country of birth, the nature of the first outcome or censoring event, the follow-up time (that is, the time from sterilisation to outcome or censoring), and a flag to indicated whether a record was censored.

4.2 Univariate exploratory data analysis

4.2.1 Sterilisation failures

The linked data set, after pruning as described in Section 3.8 above, contained 788 records for women who were recorded as having been admitted to hospital at least 28 days after their first sterilisation procedure, for reasons or procedures from which conception could be reasonably inferred. To facilitate analysis, the various ICD-9-CM and ICD-10-AM codes used to infer conception of some kind were manually grouped into four categories, with frequencies shown below in Table 7.Unfortunately the codes actually used in practice in the ISC data were, for the majority of records, insufficiently specific to be able to reliably distinguish pregnancy from delivery of a neonate (or a stillbirth).

Table 7 - Frequency of failure outcomes in study subjects

Type of outcome	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Pregnancy and/or delivery	356	45.18	356	45.18
Abortion	320	40.61	676	85.79
Ectopic pregnancy	111	14.09	787	99.87
Hydatidiform mole	1	0.13	788	100.00

There are a number of facts that can be immediately observed from these data.

The first is that the unconditional cumulative incidence of failure is 788 / 65,499 which equals 0.0122 or 1.22 failures per 100 women, with up to eight years of

follow-up (although not eight years in every case). This is comparable with the unconditional (that is, unadjusted for length of follow-up and censoring) cumulative incidence of failure reported by earlier studies as reviewed in Chapter 2. Further consideration and comparison of the conditional (also referred to as "life-table") cumulative incidence appears below. However, the fact that the "crude" failure incidence is of the expected magnitude is encouraging.

The second observation is that 14 per cent of conceptions presented as ectopic pregnancies. This is a far higher proportion of ectopic pregnancies than would be expected in the general population – the background rate of ectopic pregnancies is typically under 0.3 per cent – but it agrees well with the proportion of ectopic pregnancies amongst tubal sterilisation failures calculated in the meta-analysis McCausland in 1980: 12.3 per cent.[77]

Perhaps unsurprisingly, 40 per cent of post-sterilisation pregnancies terminated in abortion. Unfortunately the ICD codes typically used in the ISC do not permit spontaneous abortion followed by a dilatation and curettage procedure to be reliably distinguished from induced abortion. However, the literature review did not reveal any reason to suspect that tubal sterilisation results in an increased probability of spontaneous abortion, should conception subsequently occur. Therefore, it would be possible to apply age-specific proportions for spontaneous abortion to estimate the proportion of tubal sterilisation failures which are terminated by induced abortion. Such age-specific proportions for spontaneous abortion are not routinely published and estimation of them is beyond the scope of this study, however.

4.2.2 Reasons for censoring

The frequency distribution of the reasons censoring is shown in Table 8 below.

Table 8 - Frequencies of reasons for censoring in study subjects

Type of censoring	Frequency	Percent	Cumulative Frequency	Cumulative Percent
End-of-observation period	60652	93.73	60652	93.73
No longer at risk of pregnancy (hysterectomy, oophorectomy etc)	3200	4.95	63852	98.67
Post-menopausal	362	0.56	64214	99.23
Attempted reversal of tubal occlusion	280	0.43	64494	99.66
Assisted conception which bypasses fallopian tubes	163	0.25	64657	99.92
Death	54	0.08	64711	100.00

The majority of subjects were nominally observed until the end of the study period. The qualifier “nominally” is used because, as discussed in Section 3.4 above, one weakness of the passive follow-up provided by record linkage studies is that the investigator can never be certain that subjects are still effectively under observation. Nevertheless, some six per cent of study subjects were censored for reasons other than end-of-observation, which represents a 4.1 per cent reduction in person-years under observation compared to the situation if these causes of censoring were ignored. Comments on this rather modest correction to the incidence denominator as a result of record linkage (as opposed to the simple application of a cut-off date) appear in the final chapter of this work.

4.2.3 Type of sterilisation procedure

The ICD-9-CM and ICD-10-AM codes which were used to select ISC admission records representing tubal sterilisations are given in detail in Appendix C. Unfortunately, due to deficiencies in the nature of the codes and the manner in which they are typically used, little useful information about the nature of the sterilisation surgery employed can be gleaned, as evinced by Table 9 below. Each sterilisation admission contained one or more of the target ICD codes. All of the permutations of these codes which appeared in the final data set were manually allocated to semantically equivalent groups in order to create the table. Although it is clear that, unsurprisingly, the majority of procedures were performed laparoscopically, and that cautery no longer appears to be very popular, little else can be concluded from these data, particularly when, in nearly 20 per cent of admissions, the route of surgical approach is not recorded (although it is probably laparoscopic). No details about the type of salpingectomy or type of clips or banding device employed are available, and it cannot be assumed that all of the admissions in the "Sterilisation – not otherwise specified" category represent tubal banding or clipping procedures.

For these reasons, no further analysis of the type of sterilisation procedure will be undertaken because the quality of the ICD-encoded procedure data available is insufficient to warrant it. This is an unfortunate limitation of the routinely-collected administrative data used in this study.

Table 9 - Frequencies of types of tubal sterilisation surgery

Type of sterilisation	Number	Percent	Cumulative frequency	Cumulative percentage
(Bilateral) salpingectomy – laparoscopic	438	0.67	438	0.67
(Bilateral) salpingectomy – open abdominal	6	0.01	444	0.68
(Bilateral) salpingectomy - route not specified	3712	5.67	4156	6.35
Electrodestruction - route not specified	51	0.08	4207	6.42
Electrodestruction - laparoscopic	414	0.63	4621	7.06
Electrodestruction - open abdominal	1	0.00	4622	7.06
Sterilisation - not otherwise specified - laparoscopic	49146	75.03	53768	82.09
Sterilisation- not otherwise specified - open abdominal	1800	2.75	55568	84.84
Sterilisation - not otherwise specified - route not specified	9889	15.10	65457	99.94
Sterilisation - not otherwise specified - vaginal approach	42	0.06	65499	100.00

4.2.4 Age at first sterilisation procedure

The following tables and graph summarise the location and distribution of age at first tubal sterilisation procedure.

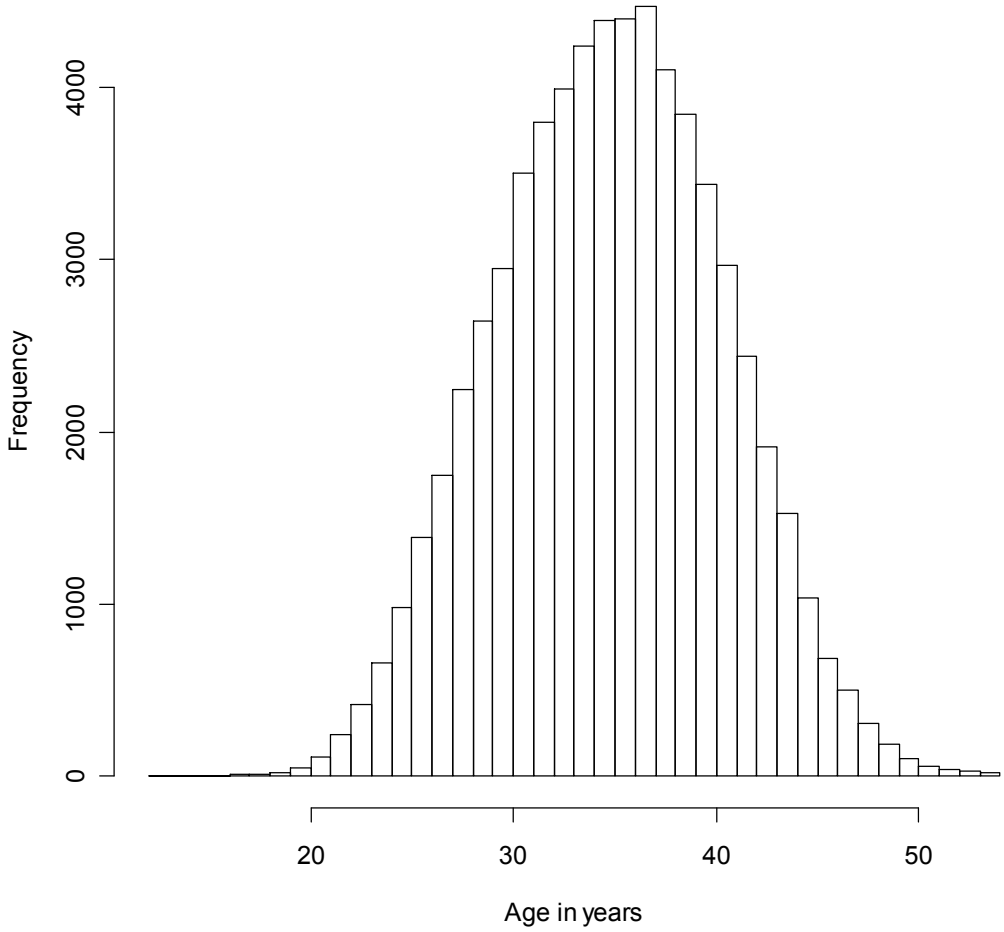
Table 10 - Summary measures of location for age in years at first tubal sterilisation procedure

Minimum	1st quartile	Median	3rd quartile	Maximum	Mean	Standard deviation
12.19	30.81	34.86	38.69	53.59	34.77	5.56

Table 11 – Frequency by age group at first tubal sterilisation procedure

Quinquennial age group	Number	Percent	Cumulative frequency	Cumulative percentage
10 - 14 yrs	5	0.01	5	0.01
15 - 19 yrs	74	0.11	79	0.12
20 - 24 yrs	2188	3.34	2267	3.46
25 - 29 yrs	10409	15.89	12676	19.35
30 - 34 yrs	19518	29.80	32194	49.15
35 - 39 yrs	20458	31.23	52652	80.39
40 - 44 yrs	10420	15.91	63072	96.29
45 - 49 yrs	2209	3.37	65281	99.67
50 - 54 yrs	210	0.32	65491	99.99
55 - 59 yrs	4	0.01	65495	99.99
60 - 64 yrs	2	0.00	65497	100.00
65 - 69 yrs	2	0.00	65499	100.00

Figure 1 - Distribution of age at first tubal sterilisation procedure

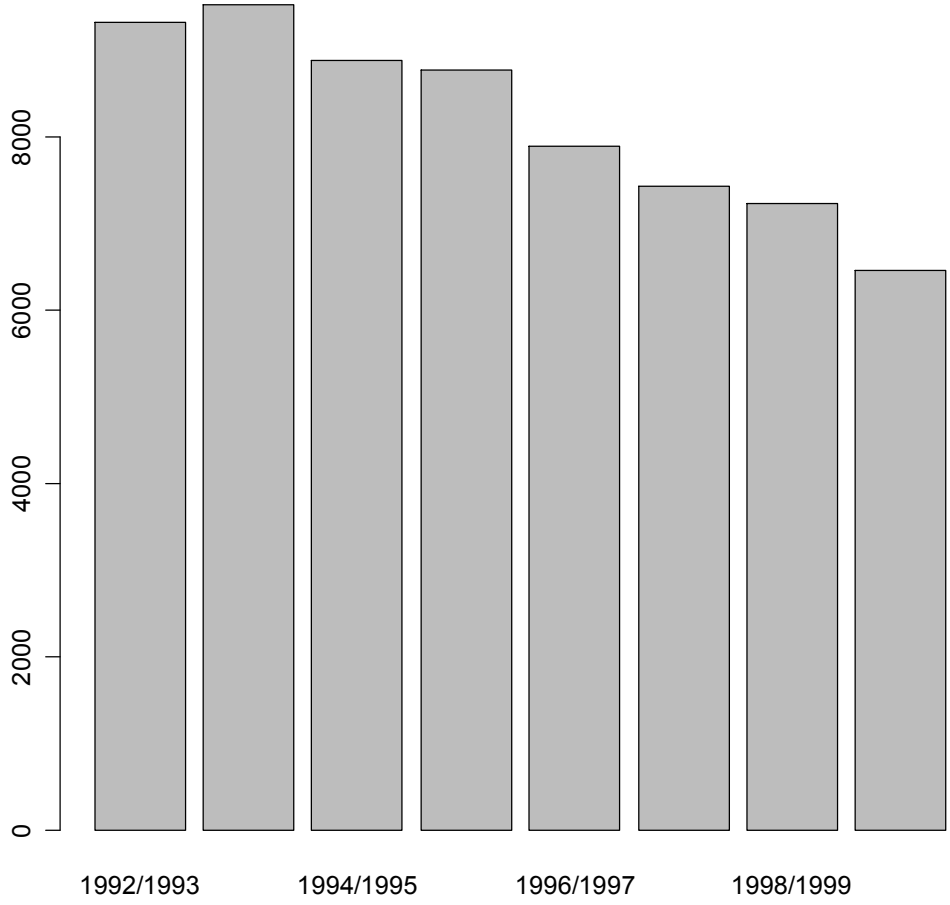


This age distribution is much as expected: surgical sterilisation is a procedure which is difficult to reverse and should be offered only to women who have completed their families. The small number of sterilisation procedures in women under the age of 20 may have been may have been performed in cases of mental handicap where the subject was thought to be unsuitable for motherhood. The procedures in women aged over 50 years are more surprising and cannot be readily explained. They may be due to salpingectomy for reasons other than contraception - a possibility which exists in younger age groups as well, although in those age groups the vast majority of salpingectomy procedures will be for the purposes of sterilisation.

4.2.5 Year of sterilisation

The numbers of hospital admissions in NSW for tubal sterilisation procedures by financial year is given in Table 6 in Section 3.6 above. The number of first-time tubal sterilisations (which is almost the same) by financial year is shown in Figure 2 below. Clearly the popularity of tubal sterilisation is falling, with the number of procedures performed in 1999/2000 some 30 per cent fewer than the number performed in 1993/1994.

Figure 2 - Number of tubal sterilisation procedures by financial year, 1992/1993 to 1999/2000



4.2.6 Health insurance and public/private election

Health insurance and private/public hospital admission payment status frequencies are shown in the following table.

Table 12 - Health insurance and private/public hospital admission payment status for NSW tubal sterilisation 1992/1993 to 1999/2000

Health insurance and private/public election	Number	Percent	Cumulative frequency	Cumulative percentage
Basic cover - private patient	2907	4.44	2907	4.44
Full cover – private patient	13399	20.46	16306	24.90
Unknown/Not stated	10168	15.52	26474	40.42
No cover - public patient	39021	59.57	65495	99.99
No cover - private patient	4	0.01	65499	100.00

In order to simplify subsequent analysis, patients were grouped into three broad categories with respect to insurance status and hospital stay payment election: private, public or unknown (24.9, 59.6 and 15.5 per cent respectively). Although both public and private hospitals have a strong financial incentive to accurately record whether private health insurance is held, because it affects the way the patient is billed for services, there is little incentive to record public patient status. Therefore it is likely that the majority of the patients in the "Unknown/Not stated" category are, in fact, public patients.

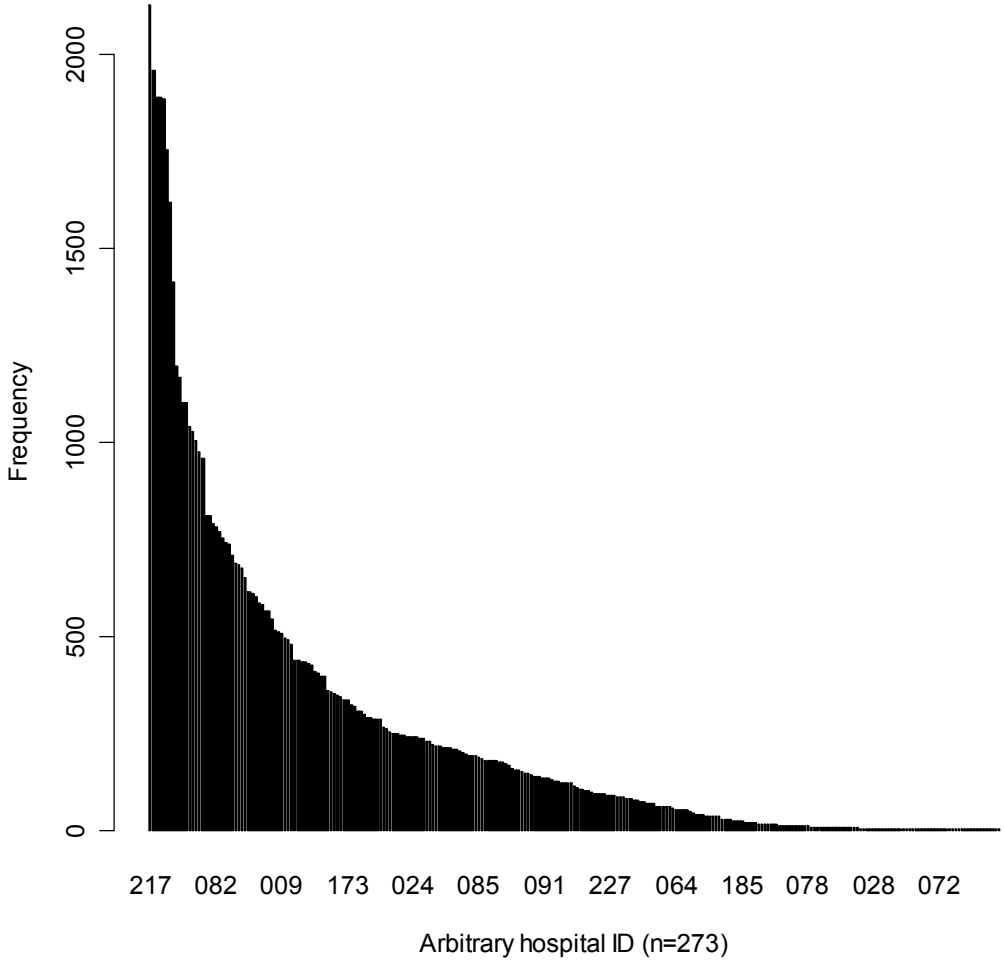
4.2.7 Hospitals performing tubal sterilisation procedures

Confidentiality restrictions prevent the identification of individual hospitals in this thesis, and prevent the inclusion of any information which could be used to infer their identity. Therefore the codes used to identify hospitals in the ISC have been replaced by randomly-chosen, arbitrary hospital identification numbers. The distribution of the number of tubal sterilisation procedures performed in each hospital over the eight year period of the study is shown in Figure 3 below. Clearly

some hospitals perform only a small number of procedures each year, and half of the procedures were performed in just 28 of the 273 hospitals represented in the data.

By analogy with other areas of health care in which failure or complication rates of procedures are inversely correlated with the total number of procedures of that type performed in a given institution each year, it is possible to hypothesise that those hospitals performing a greater number of procedures each year may have lower failure rates. Therefore, in order to facilitate exploration of this issue in subsequent analysis, the 273 hospitals were additionally grouped into tertiles based on the total number of procedures which each hospital performed over the study period, with 14, 35 and 224 hospitals in the respective quantiles.

Figure 3 - Frequency distribution of total tubal sterilisation procedures in each hospital over the eight-year study period.



4.2.8 Country of birth

For analysis purposes, countries of birth were grouped into approximately 50 groups, some representing individual countries, others representing ensembles of geographically related countries. This grouping has been widely used for population health reports such the Report of the NSW Chief Health Officer, and strikes a reasonable balance between country specificity and overly large, heterogenous aggregation of countries.[189] The frequencies of the grouped countries of birth for study subjects, up to the 95th percentile, are shown below in Table 13.

Table 13 - Frequency of countries of birth of study subjects, up to 95th percentile

Country of birth	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Australia	48030	73.33	48030	73.33
United Kingdom & Ireland	2957	4.51	50987	77.84
Lebanon	1334	2.04	52321	79.88
New Zealand	1260	1.92	53581	81.80
Philippines	1022	1.56	54603	83.36
Other	1008	1.54	55611	84.90
Vietnam	645	0.98	56256	85.89
Fiji	510	0.78	56766	86.67
Other Oceania	483	0.74	57249	87.40
Former Yugoslav Republics	473	0.72	57722	88.13
Italy	466	0.71	58188	88.84
India	361	0.55	58549	89.39
Turkey	358	0.55	58907	89.94
Other Middle East	349	0.53	59256	90.47
Austria	330	0.50	59586	90.97
Chile	324	0.49	59910	91.47
China	319	0.49	60229	91.95
Greece	298	0.45	60527	92.41
Hong Kong & Macau	288	0.44	60815	92.85
Egypt	267	0.41	61082	93.26
Portugal	266	0.41	61348	93.66
Germany	240	0.37	61588	94.03
South Africa	229	0.35	61817	94.38
Malaysia & Brunei	218	0.33	62035	94.71
Indonesia	198	0.30	62233	95.01
Malta	181	0.28	62414	95.29
Other South America	181	0.28	62595	95.57
Sri Lanka	181	0.28	62776	95.84

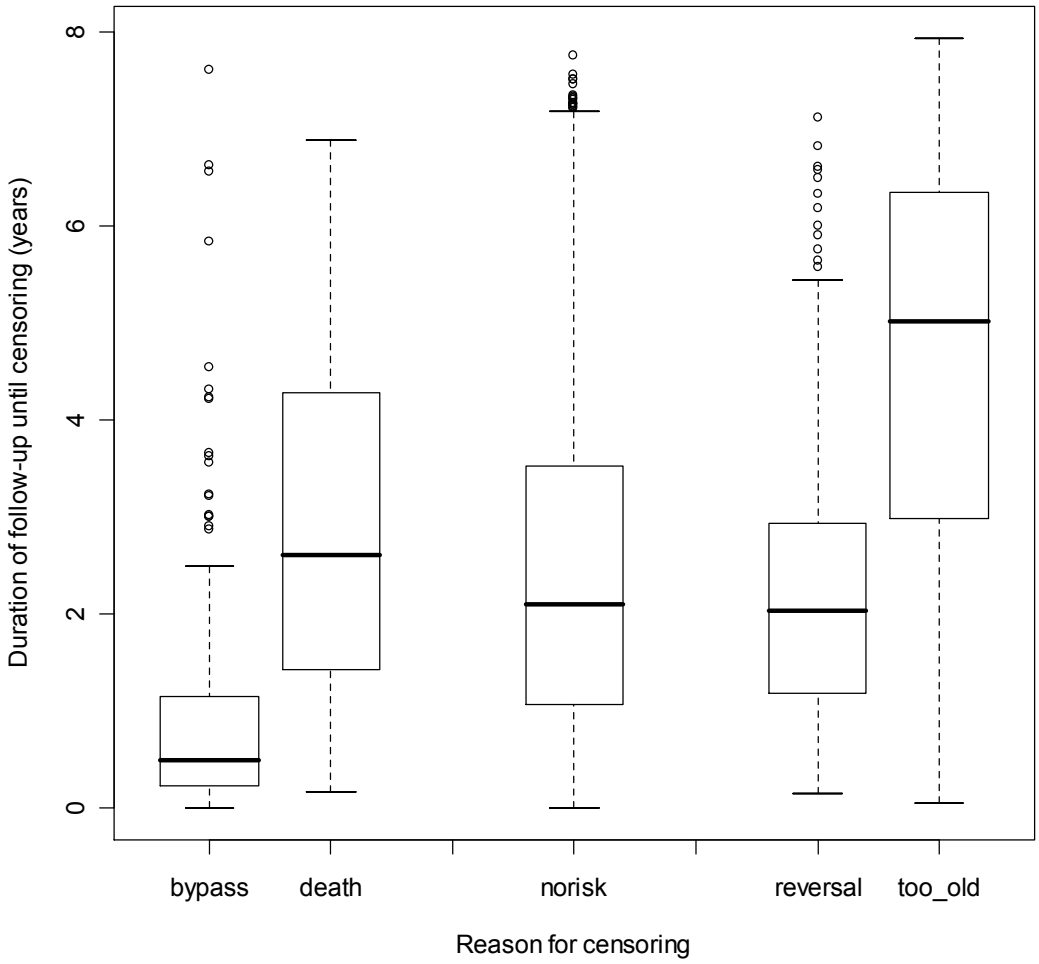
4.3 Selected bivariate contrasts

The relationships between all of the variables and variates in the final dataset were examined using typical exploratory analysis techniques.[190] Only those of interest will be presented here. Of course, the relationships of greatest interest are those between the covariates and the outcome variables, which are the combination of fact-of-failure and time-to-failure. However, these relationships will be explored in the next section using plots of Kaplan-Meier estimators and nonparametric tests to assess the statistical significance of variation in outcomes by covariate, where such heterogeneity is observed.

4.3.1 Duration of follow-up until censoring for censored subjects

Although time to failure is closely scrutinised in most survival analyses, time-to-censoring is often not examined. The extrema, quartiles and median duration of follow-up until censoring for each of the reasons for censoring are shown in the Box plots in Figure 6 below. The median time to first reversal attempt was just over two years. Surprisingly, the median time to attempted assisted reproduction ("bypass" of Fallopian tubes) was under six months. These records were reviewed to exclude data errors or false links, but no anomalies were detected. This is an unexplained observation, but one which will not be pursued further here as it is outside the scope of this study.

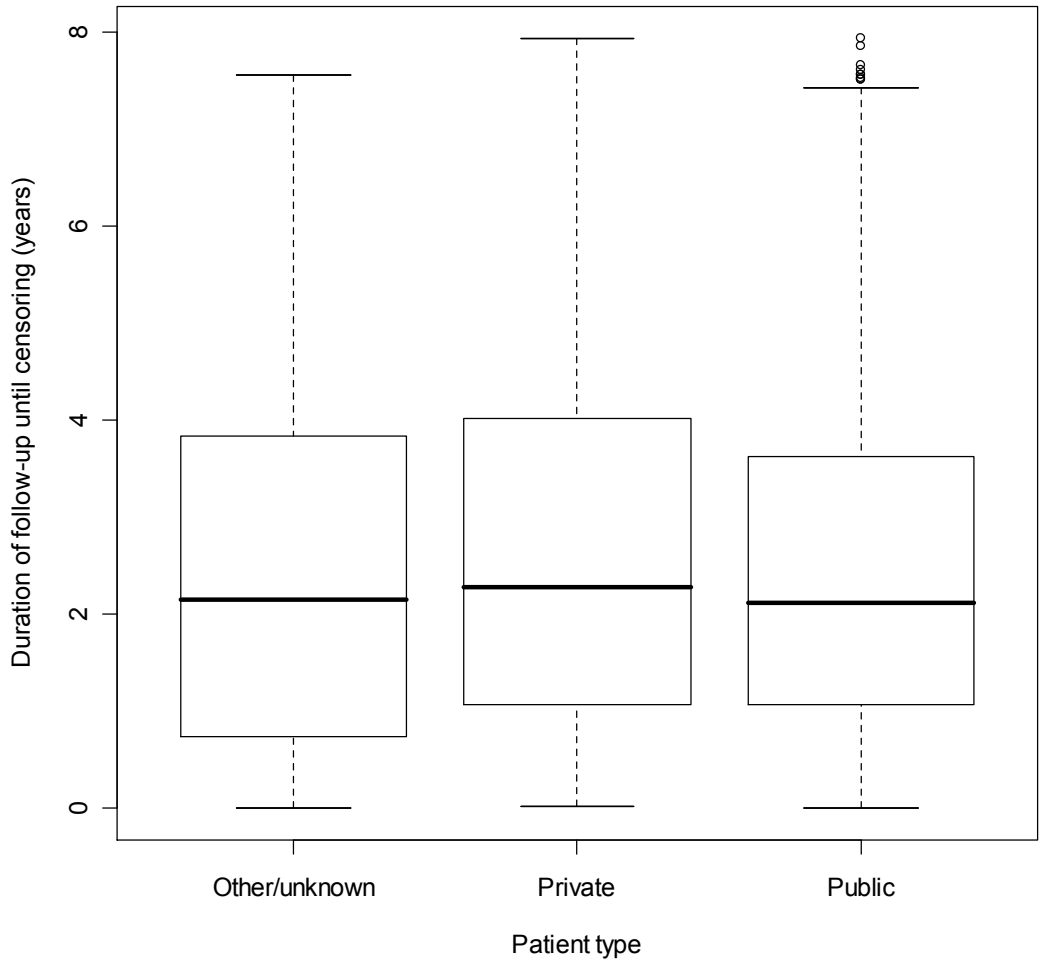
Figure 4 - Duration of follow-up until censoring, by reason for censoring



As mentioned earlier, it is important that censoring in survival analyses is not associated with covariates - at least in those analyses using life-table, Kaplan-Meier estimators and ordinary proportional hazards or parametric survival modelling techniques – some more advanced methods can accommodate conditional, non-random censoring. Using Box plots, such as the one shown below in Figure 5, association with length of follow-up was visually assessed for each of the covariates in the data set (Student's t-tests or analysis of variance could also have been used). Apart from age at sterilisation, no association with length of follow-up was detected. The association between age at sterilisation and length of follow-up is expected and unavoidable, because older subjects are necessarily censored due to menopause

sooner (and thus more often) than younger subjects, and also have a higher risk of undergoing hysterectomy/oophorectomy than younger subjects. However, it is unlikely that significant bias will have been introduced due to these associations, given that censoring other than for end-of-follow-up accounted for a reduction of denominator time of just 4.1 per cent, as noted above.

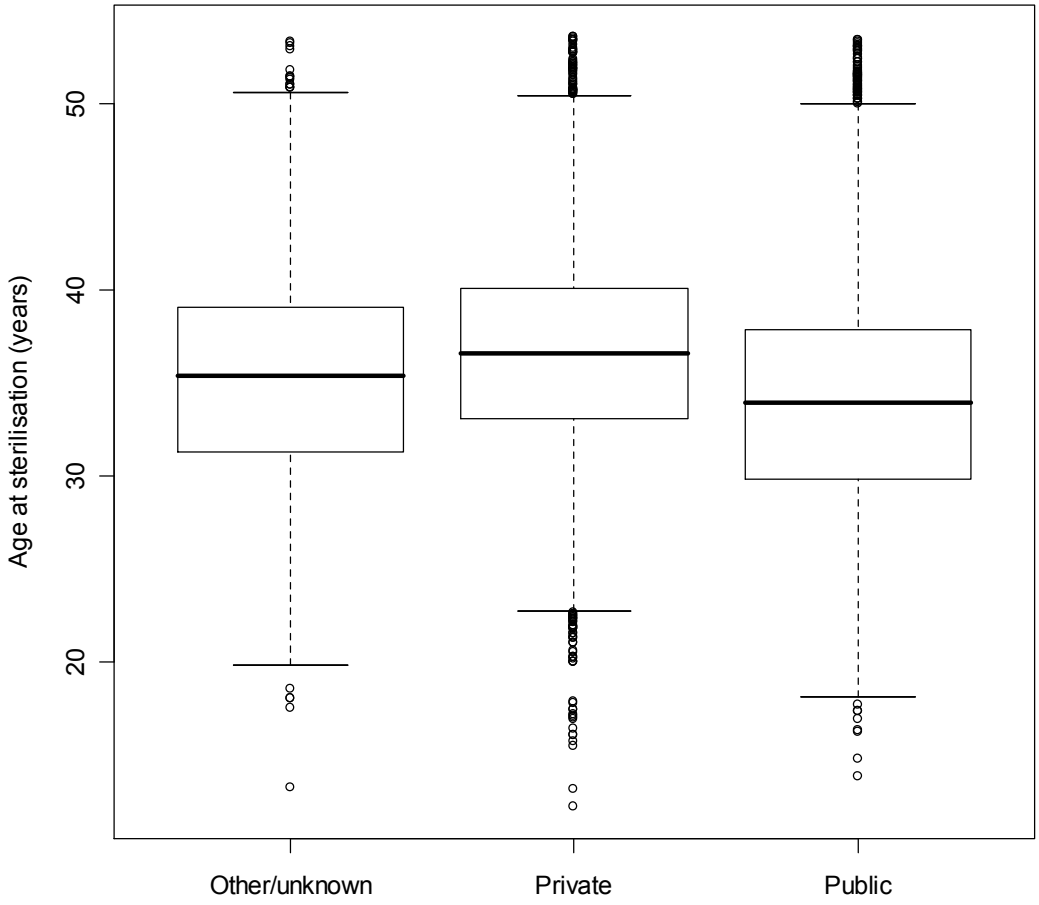
Figure 5 - Duration of follow-up until censoring, by private/public patient status



4.3.2 Age by private/public patient status

As shown in the Box plots in Figure 6 below, private patients were slightly older on average than public patients, but the difference in age distributions between the groups was not marked.

Figure 6 - Box plot of age at sterilisation by private/public patient status at sterilisation admission

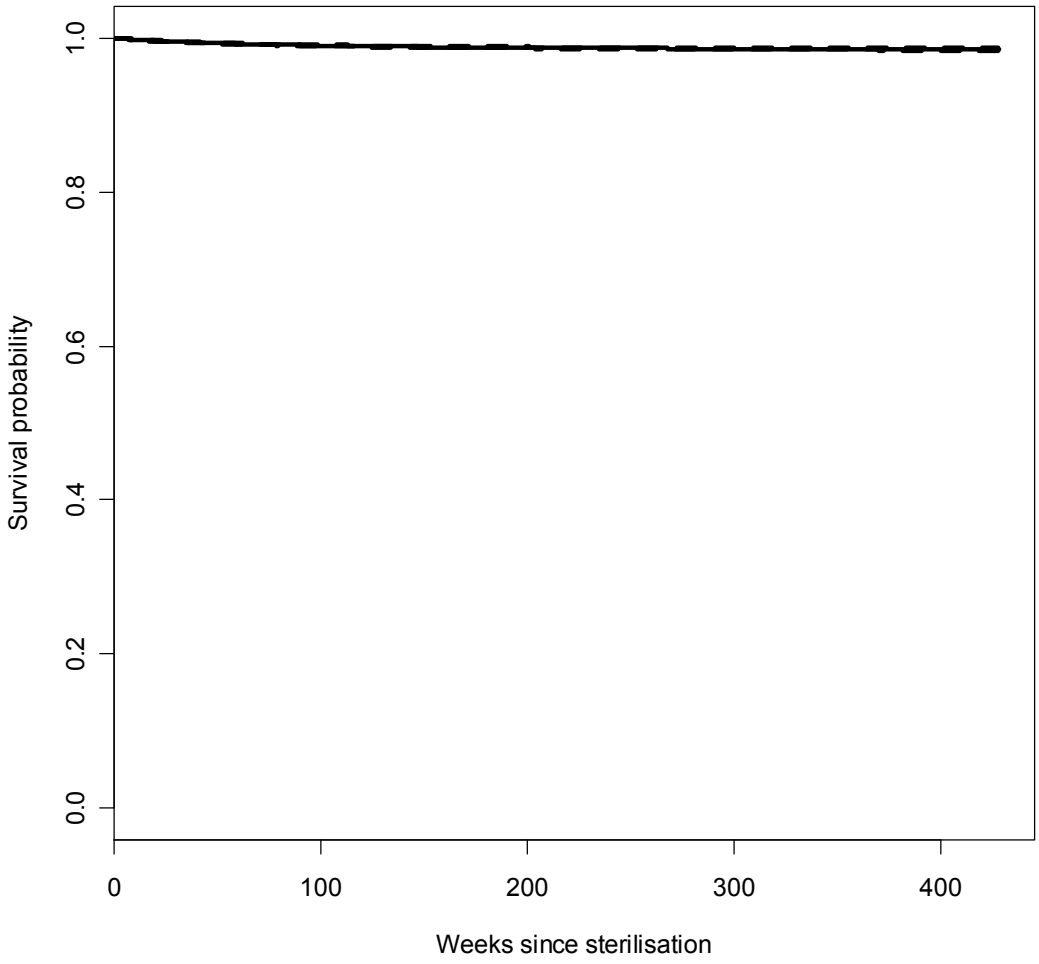


4.4 Kaplan-Meier analysis

Unless otherwise noted, all survival analyses were undertaken using the `survival` library by Therneau, originally written for the S/Plus statistical environment and ported to the R statistical package by Lumley.[191]

The Kaplan-Meier survival curve for the entire study data set is shown in Figure 7 below.

Figure 7 - Kaplan-Meier survival (non-failure) probability curve for all study subjects



However, due to the low incidence of failure (in absolute terms), the shape of the curve is difficult to appreciate. This situation is remedied by applying a log transformation to the probability of survival and adjusting the graph axis, as shown in Figure 8 below –. The instantaneous hazard function, estimated using the Mueller and Wang kernel-based methods as implemented in the `muhaz` library for R, is shown in Figure 9.[192]

Figure 8 - Kaplan-Meier survival (non-failure) probability curve, with Goodman 95% confidence intervals, for all study subjects

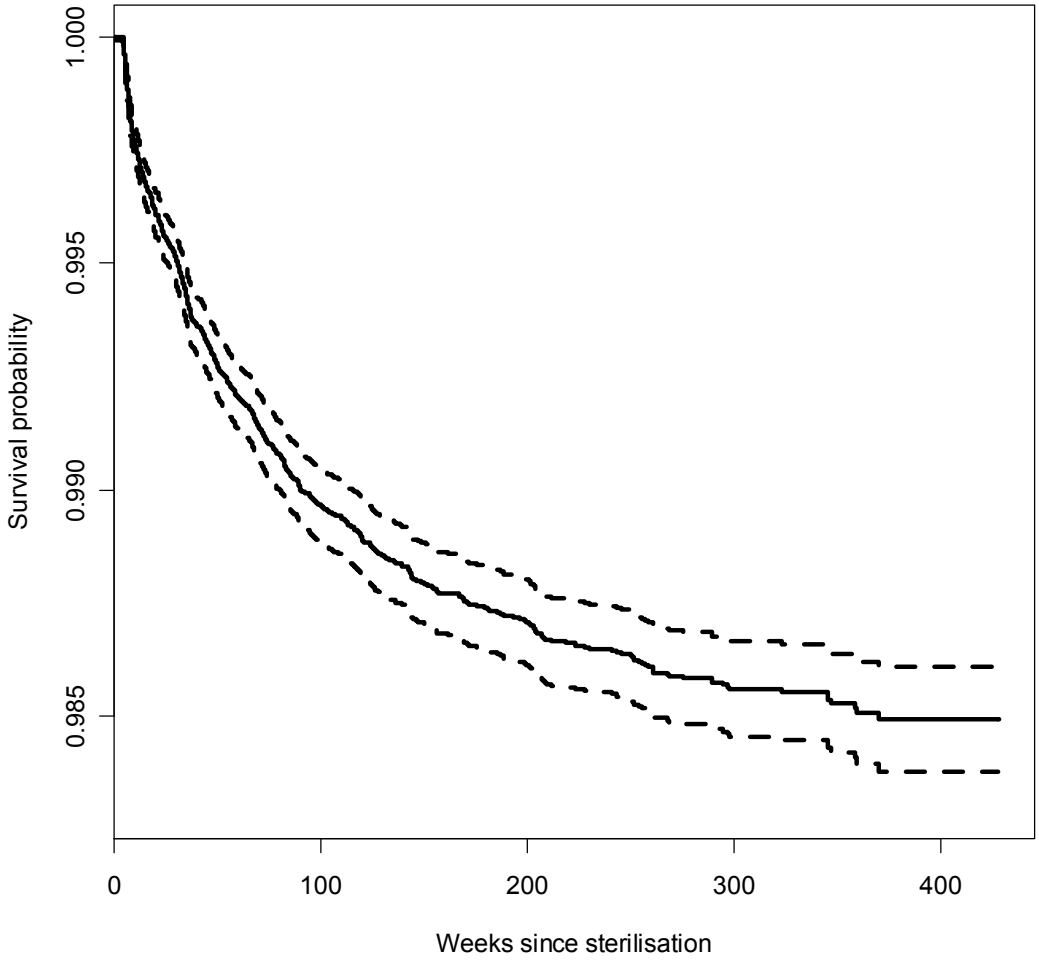
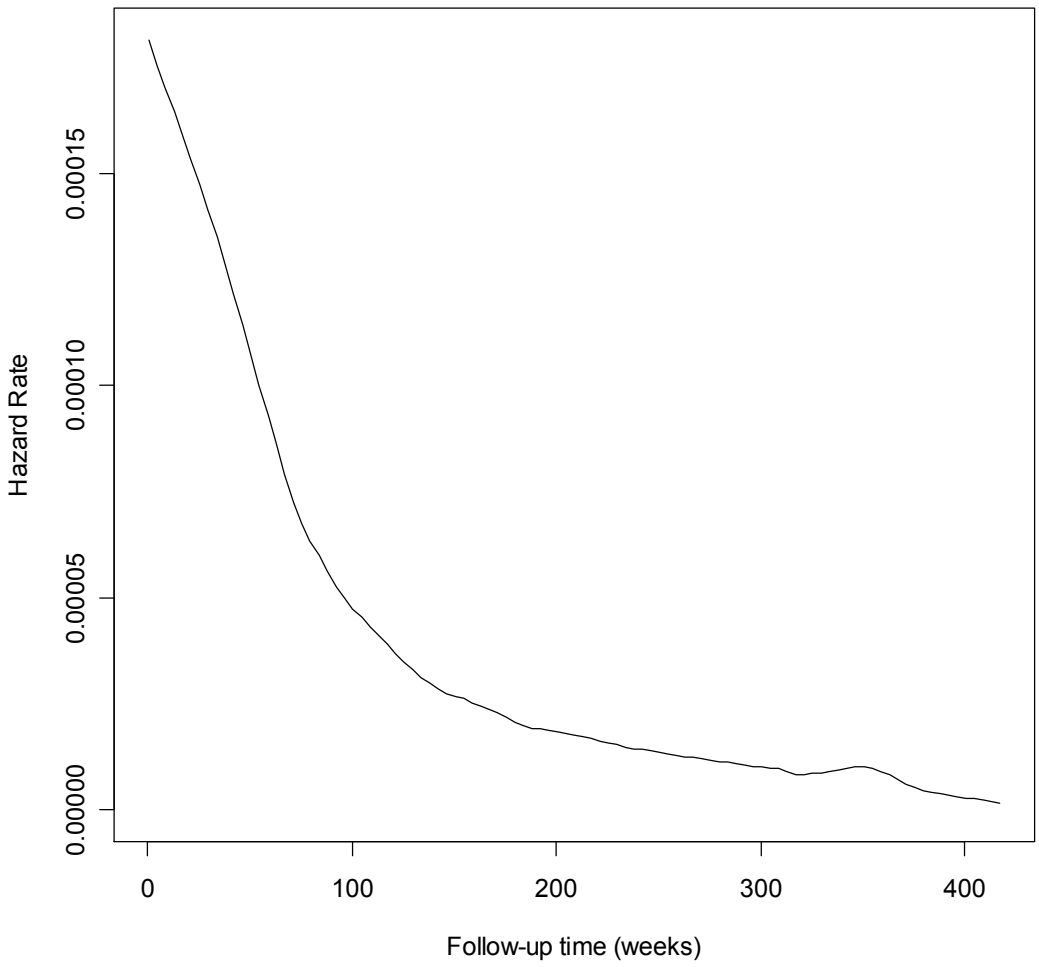


Figure 9 - Mueller and Wang kernel-based smoothed estimate of the hazard rate for all study subjects



Several observations can be made about these plots. The first is that the survival probability and hazard rate curves are gratifyingly smooth and "biological" in appearance – admittedly the Mueller and Wang nonparametric hazard rate curve is smoothed by nature of its kernel-based estimation, but the smoothing parameters used were chosen so that only minor variations are suppressed. Only a small bump in the hazard rate at about 380 weeks of follow-up mars the curve. It is quite likely that data errors or false links would show up as "artefacts" in the instantaneous hazard and the resulting survival function.

The second observation is that the shape of the hazard function mirrors that reported by Peterson *et al*.[65] in the long-term CREST study – the hazard of failure is greatest just after the sterilisation procedure, with an exponential decline over the following years, but with some residual hazard of conception even after eight or more years.

Thirdly, the conditional (Kaplan-Meier) cumulative incidences of failure at annual durations of follow-up, as given in Table 14 below, are very comparable to those reported by the two largest and most recent studies of tubal sterilisation located as part of the literature review in Chapter 2 above, as shown in Figure 10 below. In order to illustrate the equivalence of actuarial life-table calculations of survivorship and Kaplan-Meier estimators, both values are shown in the table – the very small differences are due to the greater precision of the Kaplan-Meier method due to its use of the finest possible time increments during calculation.

Although the approximate correspondence between the shapes of the survival curves and the magnitudes of the failure rates in this and the other two large studies is gratifying and re-assuring, it is nevertheless reasonable to ask whether such correspondence to be expected, or has it occurred purely by chance? Such a question, which is also at the heart of many meta-analyses, is essentially philosophical in nature and cannot readily be decided empirically, short of performing more studies.

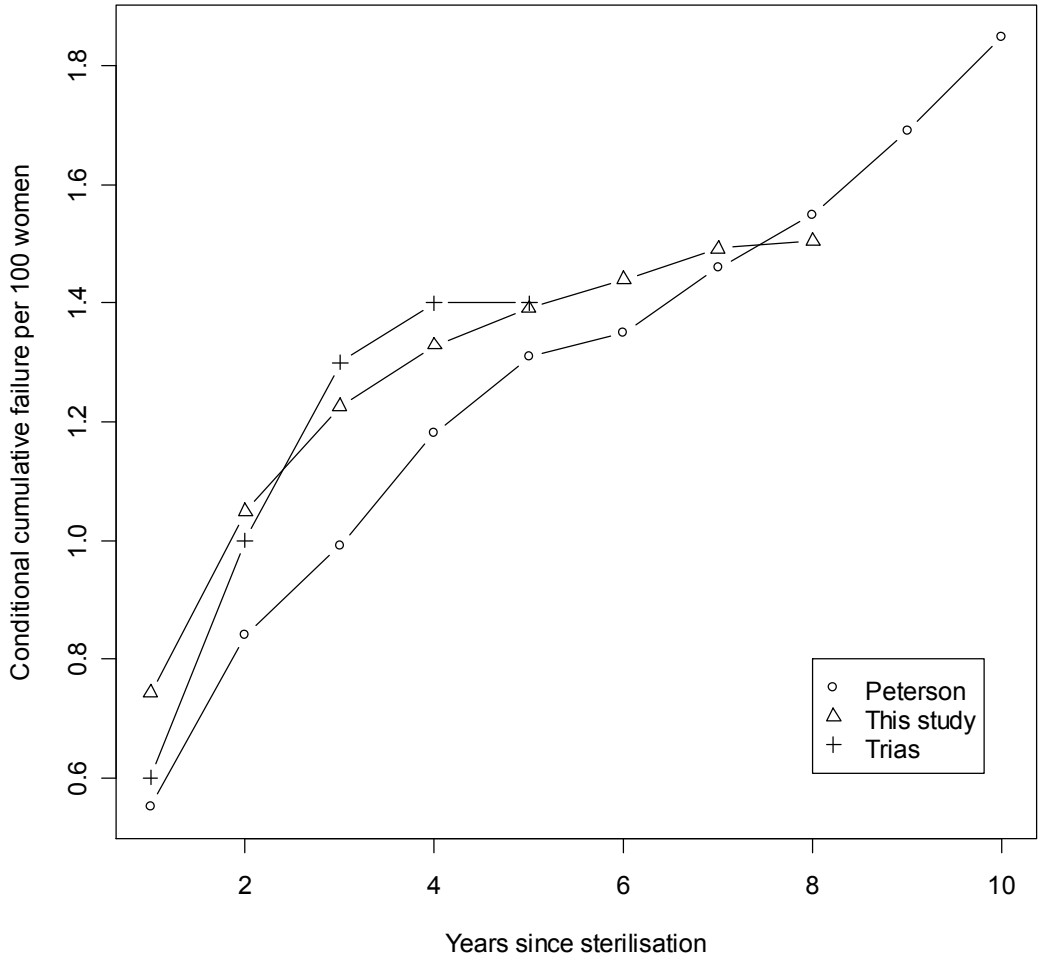
From a stochastic point of view, none of the designs of the three studies in question in any way constrained the value of the cumulative failure rate to a particular range of values which were *a priori* considered to be feasible– each of the studies might well have produced much higher or much lower failure rates. Thus, it would seem to be quite unlikely that three studies, each using fairly large numbers of subjects,

would each produce estimated cumulative failure rates which were quite close to each purely by chance. From an epidemiological perspective, the question might be decomposed into three components: a) whether the internal validity of each of the three studies is such that they accurately estimate the actual failure rate in their respective study populations; b) whether the external validity of each of the studies is such that their results can be generalized to wider communities and time periods in the countries which the studies were carried out; and c) whether fundamental differences in failure rates might be expected between countries and/or time periods. The CREST study probably has the greatest internal validity due to its prospective design, the resulting extensive information on potential confounders available to the investigators, and the thoroughness and duration of follow-up. The study reported here may have the least internal validity, due to inevitable biases towards the null (or towards underestimation of the failure rate) due to limitations in the follow-up process through record linkage, as discussed in detail in Section 3.4 of this work. In terms of external validity, this study, being essentially population-based rather than clinic-based as the other two are, undoubtedly has the best generalisabilty. However, both internal validity and external validity cannot be quantified, only considered and hypothesized about. That leaves the issue of whether one would expect the true, underlying failure rates of female surgical sterilisation to differ significantly between populations in the US and in a relatively wealthy Latin American country in the 1980s from those in NSW in the 1990s. Again, this is a matter for debate, but in the author's opinion, there is no reason to expect the underlying failure rates to differ significantly, because the surgical techniques used and the technical competency of those applying those techniques are likely to be broadly comparable between all three countries and study periods.

Table 14 - Life-table (product-limit) estimates from this study of cumulative incidence of failure at annual intervals post-sterilisation

Years since sterilisation	Number still at risk	Number of events	Cumulative incidence of failure per 100 women (actuarial life-table)	Cumulative incidence of failure per 100 women (Kaplan-Meier estimator)	Lower 95% CI (Kaplan-Meier estimator)	Upper 95% CI (Kaplan-Meier estimator)
1	57700	464	0.75	0.74	0.68	0.81
2	49637	167	1.06	1.05	0.97	1.13
3	41778	82	1.24	1.23	1.14	1.32
4	33850	39	1.34	1.33	1.23	1.42
5	25401	18	1.40	1.39	1.29	1.49
6	17065	11	1.45	1.44	1.34	1.54
7	8379	6	1.50	1.49	1.38	1.60
8	45	1	1.52	1.51	1.39	1.62

Figure 10 - Conditional (life-table) cumulative incidence of tubal sterilisation failure per 100 women as reported by Peterson *et al.*[65], Trias *et al.*[63] and in this study.



Finally, the 95 per cent confidence intervals around the Kaplan-Meier estimators for this study are reasonably narrow, despite the relatively rare nature of the failure events being studied. In other words, the precision of the study is quite good. This is due to the fact that it is population-based and that that population base is reasonably large (between six and seven million people). Indeed, the number of subjects in this study is approximately twice as many as in the largest published study to date, and over six times as many as in the most expensive study to date (the CREST study). This illustrates the value of record linkage techniques applied to routinely collected health data in bringing considerable statistical precision (and power) to the study of even rare events, despite the inherent deficiencies and limitations of the source data.

4.5 Incidence of failure by covariates

The subsequent sections of this Chapter examine the available covariates with respect to time-to-failure. Kaplan-Meier survival curves with 95 per cent Greenwood confidence limits (with the Kalbfleish and Prentice transformation to constrain them to the interval 0 to 1) were used to visualise differences in failure incidence. The `survdiff` function in the R `survival` library, which implements the G-rho family of tests as described by Harrington and Fleming was used to provide non-parametric tests of differences between the survival curves.[193] Both the log-rank test (also known as the Mantel-Haenszel test) and Gehan-Wilcoxon score test were used but only the log-rank test results are reported because both tests gave almost identical results in all cases.

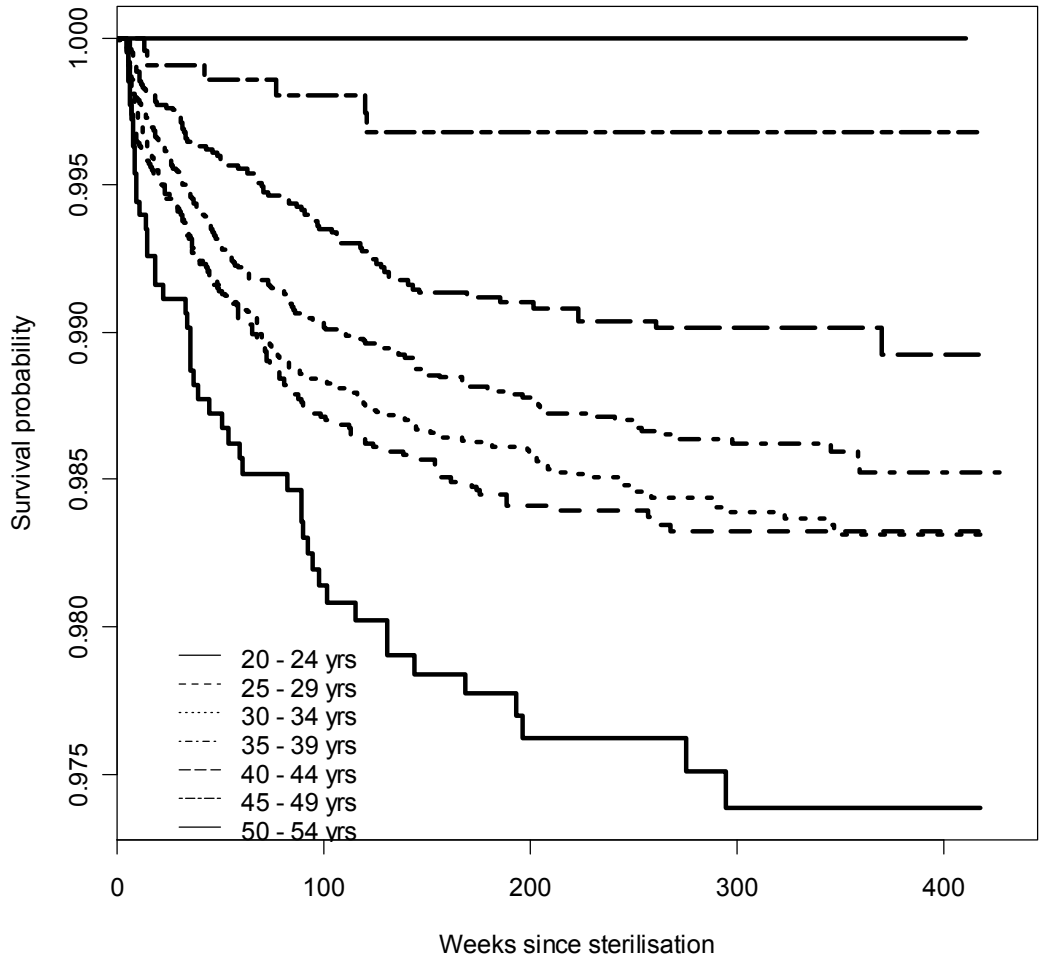
4.5.1 Incidence of failure by age group

Survival (non-failure) curves for each age group are shown in Figure 11 below. It should be noted that due to difficulties in formatting plot legends in the R statistical package, the order of the legend is the reverse of the order of the survival curves. In other words, the survival curve for the oldest age group is the uppermost line on the graph, the curve for the second oldest age group is the second line from the top on the graph and so on. The survival curves fall, in fact, in exact age group order, which indicates that age is inversely associated with failure of tubal sterilisation. The log-rank test gives a chi square value of 57 on 11 degrees of freedom, which is highly significant.

Peterson *et al.*[65] also found this association of failure with age, reporting an adjusted relative risk 1.25 for 18 to 27 year olds and 0.46 for 34 to 44 year olds relative to 28

to 33 year olds. This age trend is almost certainly related to the fact that fertility of both women and men is known to decline with age, as is frequency of sexual intercourse.

Figure 11 - Kaplan-Meier survival (non-failure) probability curve, by age group

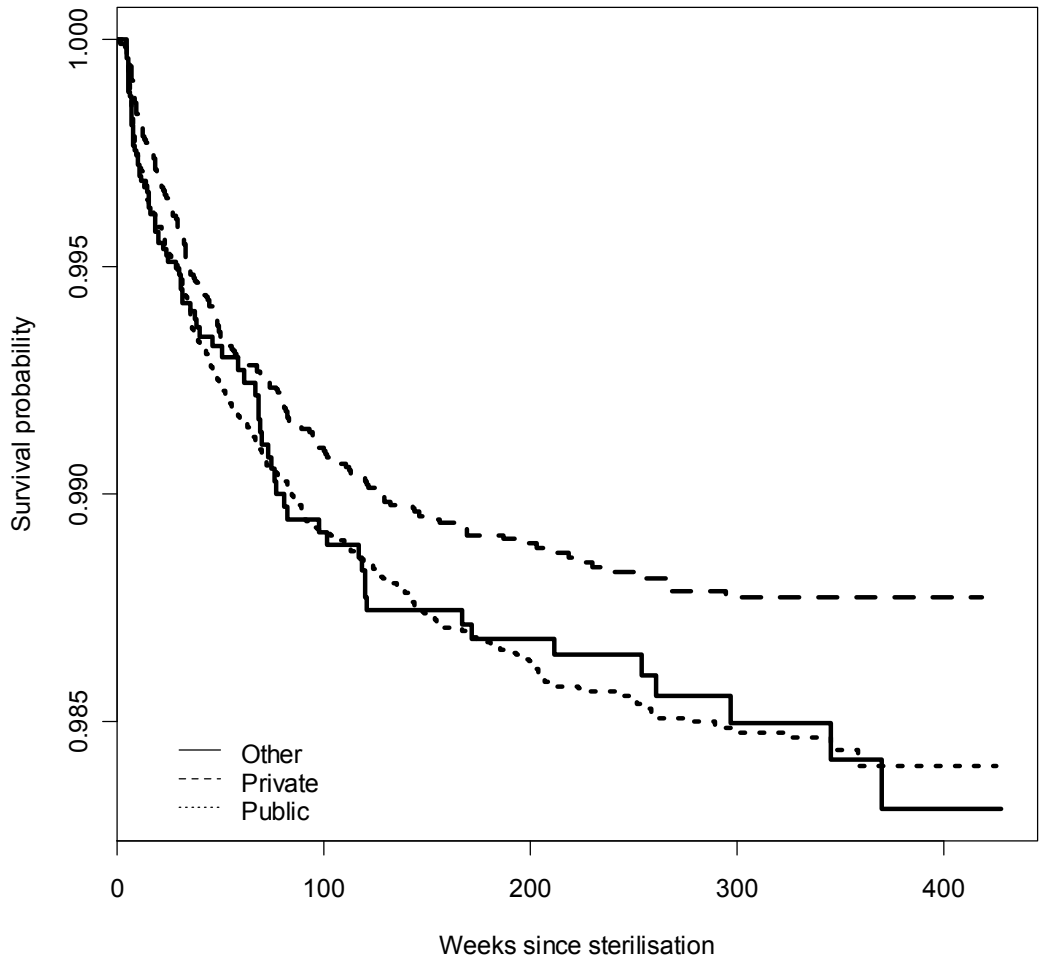


4.5.2 Incidence of failure by private/public patient status

Survival (non-failure) curves by private/public patient status are shown in Figure 12 below. The log-rank test gives a chi square value of 6.6 on two degrees of freedom, which is equates to a (arbitrarily significant) p-value of 0.0364. The survivorship functions for the public and other/unknown groups are very similar, which reinforces

the presumption mentioned in Section 4.2.6 above that patients with other or unknown status are, in fact, public patients.

Figure 12 - Kaplan-Meier survival (non-failure) probability curve, by private/public patient status at time of first sterilisation procedure



We have seen in Section 4.3.2 that private patients tend to be slightly older, on average, than public patients, and we have noted in Section 4.5.1 that the incidence of failure is inversely related to age. Therefore, it is possible that at least some of the apparent difference in failure incidence between public and private patients is due to confounding by age differences. Private patients are also far more likely to be treated in private hospitals, which tend to be smaller institutions, and as we shall see in the next section, the size of institution is also related to incidence of failure – smaller hospitals tend to have lower rates of sterilisation failure. This may also be

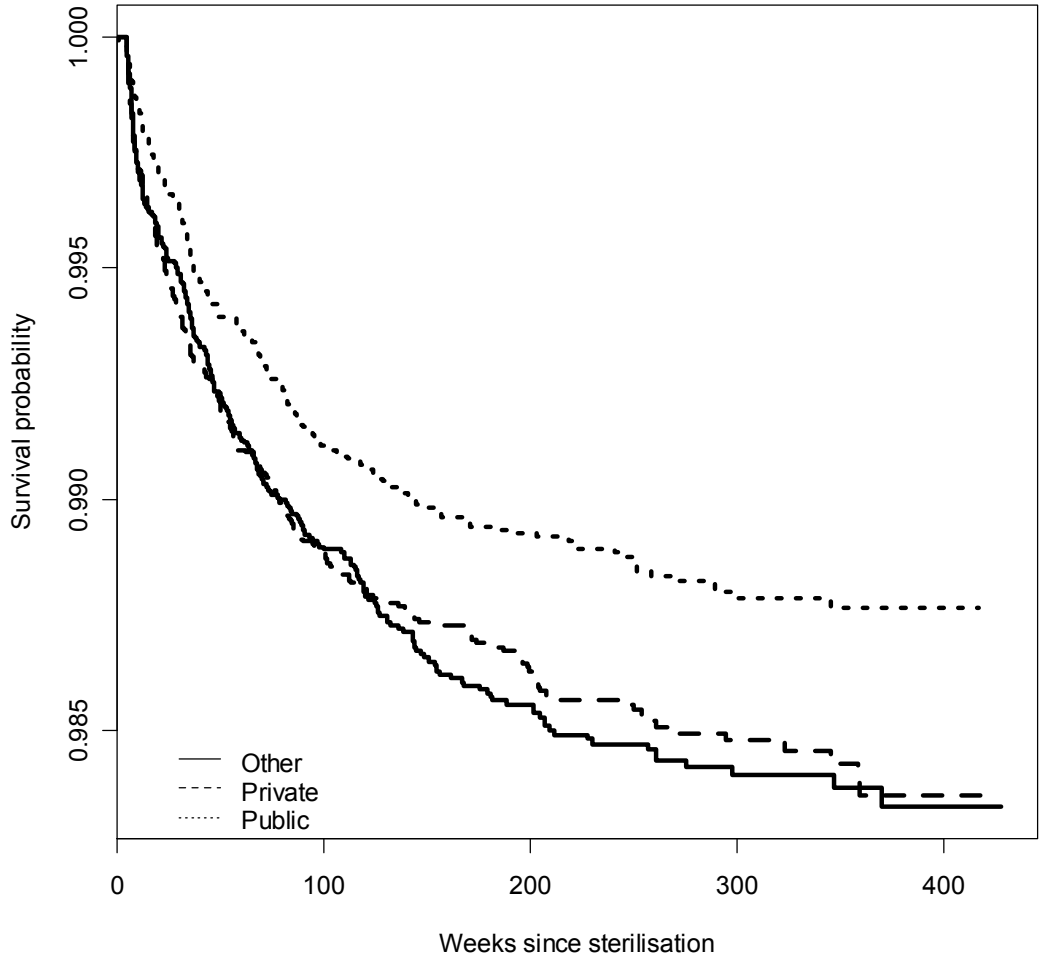
confounding the apparent association with public/private patient status. However, there may also be a residual association, even after adjusting for such confounding – this will be explored further in Chapter 5. None of the published studies reviewed examined differences in failure incidence between publicly-funded versus privately-funded tubal sterilisation procedures, either at a clinic or hospital level or at an individual patient level, as here.

4.5.3 Incidence of failure by hospital tubal sterilisation workload tertile

Survival (non-failure) curves by private/public patient status are shown in Figure 12 below. The log-rank test gives a chi square value of 12.1 on two degrees of freedom, which is corresponds to a p-value of 0.00233, indicating statistically significant heterogeneity in failure incidence.

Of great interest is the fact that the third tertile, comprising smaller hospitals, or least hospitals which perform fewer tubal sterilisations procedure each year, has, on average, a substantially lower failure rate than the two upper tertiles of hospitals that perform more procedures. As noted above, private hospitals tend to be smaller, and private hospitals treat far more privately insured patients, and thus this association may also be confounded.

Figure 13 - Kaplan-Meier survival (non-failure) probability curve, by tertile of hospital workload of tubal sterilisation procedures



It is possible to speculate about other reasons for this inverse association between tertile of institutional tubal sterilisation workload and incidence of failure. The CREST study found statistically significant heterogeneity in failure incidence in only one of the institutions participating in the study (hazard ratio 2.06, 95% CI 1.05 to 4.04), statistically non-significant hazard ratios of 0.55 to 3.46 were found in other participating hospitals. All institutions in the CREST study were teaching hospitals, and the authors wondered whether the training carried out in these institutions resulted in higher failure rates, particularly if the training were inadequately supervised. Stoval *et al*. reported histological evidence (after subsequent partial salpingectomy) of incorrect application of tubal clips and bands in women who

became pregnant after initial tubal sterilisation as part of a residency training program for gynaecologists.[60]

4.5.4 Incidence of failure by individual hospital

Given the differences in failure between hospital tertiles, evidence of heterogeneity of failure between individual hospitals is of considerable interest. There are too many hospitals to show such differences in a plot, but the log-rank test returns a chi quare of 345 on 258 degrees of freedom, with a corresponding p-value of 0.000223. Confidentiality restrictions prevent exhaustive reporting of the observed and expected frequencies of failure in each hospital, but results for two hospitals which displayed markedly elevated incidence of failures are shown in Table 15 below.

Table 15 - Numbers of observed and expected failures in selected obfuscated hospitals

Obfuscated hospital ID	Number of tubal sterilisation procedures over study period	Observed number of failures	Expected number of failures	$(O-E)^2/E$	$(O-E)^2/V$
271	686	28	8.4	4.60	46.5
038	1618	33	19.1	1.01	10.3

An alternative method of further analysis would involve the fitting of a mixed-effects "frailty" survival model (also known as a "multilevel" model), in which a random effect at the hospital level is included in the model, allowing the estimation of a relative risk of failure between hospitals after adjusting for fixed effects (of age, insurance status and other covariates) at the individual level. Such an analysis was not performed for this study.

The failures in these hospitals can be further examined by year (Table 16), which shows that in hospital 271 the majority of failures occurred in women sterilised there

during a single financial year (1993/94), while in hospital 038 an excess of the sterilisations performed in 1996/97 and 1997/98 failed. These temporal patterns strongly suggest some relatively transient factor, such as equipment failure or the temporary employment of a registrar or surgeon with particularly bad technique. Further examination of these apparently "special cause" clusters of failure is beyond the scope of this study. Nevertheless, deletion of these failures from the data does not materially affect the overall estimated incidence of failure. It may also be argued, cogently, that their inclusion better reflects real-world efficacy of tubal sterilisation techniques, although paradoxically it also reduces the external validity or generalisability of the results.

Table 16 - Numbers of tubal sterilisation failures by financial year in two selected hospitals

Financial Year of sterilisation	Obfuscated hospital of sterilisation	
	038	271
1992/1993	3	3
1993/1994	4	17
1994/1995	4	2
1995/1996	3	1
1996/1997	8	3
1997/1998	7	1
1998/1999	3	1
1999/2000	1	0
Total	33	28

4.5.5 Incidence of failure by grouped countries of birth

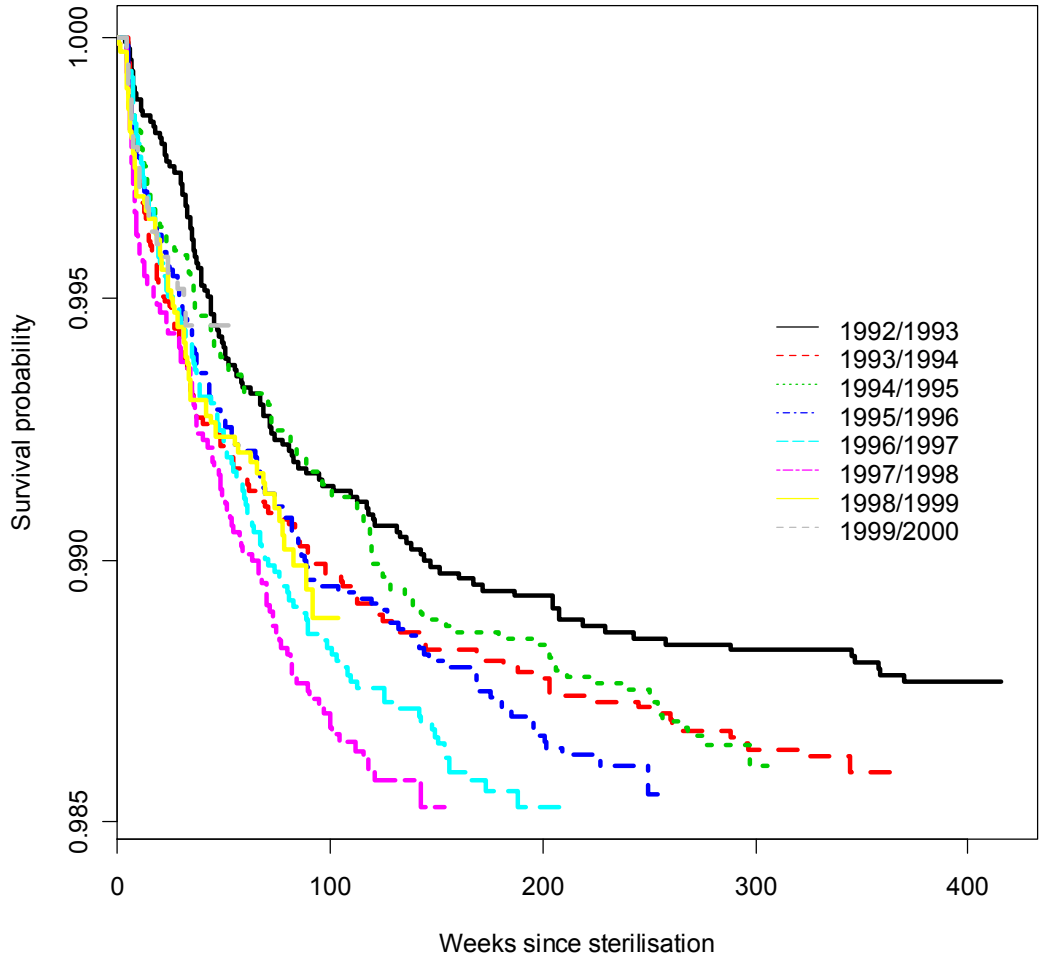
Although the log-rank test for heterogeneity in survivorship between countries of birth was statistically significant (chi square 1802 on 71 degrees of freedom, p=

0.0000008), none of the individual countries or groups of countries demonstrated markedly unexpected failure rates. In other words, the statistical test for differences was positive due to the large number of subjects, but no particular patterns or features in these differences were evident. Country of birth was not examined further in this study, although it too could be analysed using a mixed-effects frailty model, with random effects for the country of birth.

4.5.6 Incidence of failure by financial year in which sterilisation was performed.

Survival (non-failure) curves for each financial year in which the sterilisation procedure was performed are shown in Figure 14 below. The log-rank test gives a chi square value of 13.2 on seven degrees of freedom, p= 0.0681.

Figure 14 - Kaplan-Meier survival (non-failure) probability curve, by financial year in which tubal sterilisation procedure was performed



Although the log-rank test suggests that there may be hetereogenity in failure rates between years, it does not provide strong evidence for it. However, tabulation of observed and expected numbers of failures for each year is instructive - Table 17 clearly demonstrates that 1996/97 and 1997/98 were particularly bad years with respect to tubal sterilisation failures in NSW. Again, one may speculate that there

may be some special cause for this, such as the introduction of new equipment or a change in protocols or procedures.

Table 17 - Observed and expected numbers of sterilisation failures by financial year in which sterilisation was performed

Financial year in which sterilisation was performed	Number of subjects	Observed number of failures	Expected number of failures	$(O-E)^2/E$	$(O-E)^2/V$
1992/1993	9320	112	137.1	4.606	5.783
1993/1994	9509	129	135.5	0.315	0.385
1994/1995	8871	117	123.4	0.329	0.394
1995/1996	8768	120	116.5	0.104	0.123
1996/1997	7877	112	98.3	1.915	2.212
1997/1998	7430	104	83.5	5.010	5.698
1998/1999	7222	68	65.4	0.104	0.116
1999/2000	6451	26	28.2	0.173	0.187

CHAPTER 5

Proportional-hazards analysis

In Chapter 4 a number of univariate associations between covariates and the survival function were found, and the possibility of confounding and/or effect modification between at least some of these covariates was mooted. In order to simultaneously control for such confounding, it is necessary to use multivariate modelling techniques. For time-to-failure data, the proportional hazards model is the most widely used, primarily because it offers the flexibility of generalised linear models and does not assume any particular distribution for survival times.

5.1 Cox proportional-hazards regression

An exhaustive exposition on the theoretical basis for Cox regression would be redundant – there are excellent accounts in most statistical modelling textbooks. Only a very brief outline of the salient characteristics of the proportional-hazards model will be presented. This account has been synthesised from the treatments of Cox regression by Hosmer and Lemeshow, Harrell and, in particular, Fox.[23,31,194]

T_i represents the time to failure or censoring in each individual (the survival time) and t an arbitrary time during a study. It is assumed that T is a random variable, with a cumulative distribution function $P(t) = Pr(T \leq t)$ and probability density $p(t) = dP(t)/dt$. The survival function $S(t)$ is the complement of this distribution function, $S(t) = Pr(T > t) = 1 - P(t)$. The hazard function, $h(t)$ is the instantaneous risk of failure at time t, conditional on survival to that time.

Statistical models of survival data are fit to the hazard function or its logarithmic transformation. For example, if a constant hazard over time, $h(t) = \nu$, is assumed then an exponential distribution of survival times is implied (analogous to, say, the rate at which compound interest accrues), with density function $p(t) = \nu e^{-\nu t}$.

The goal of survival analysis modelling is to examine the relationship of the survival distribution to covariates, and this typically entails the specification of a linear model for the log of the hazard function. For example, a parametric model based on the exponential distribution may be written as:

$$\log h_i(t) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}$$

or, equivalently:

$$h_i(t) = \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik})$$

where the subscript i denotes each subject, k each covariate and α denotes a baseline hazard.

The advance offered by the Cox model is that it leaves the baseline hazard function $\alpha(t) = \log h_0(t)$ unspecified:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}$$

or equivalently:

$$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik})$$

It should be noted that this model is *semi-parametric* because, although the baseline hazard function can take any form, the covariates must have a linear relationship to the instantaneous hazard of failure. The real breakthrough made by Cox in 1972 was the realisation that this model can still be estimated, despite the unspecified baseline hazard function, using the method of partial likelihood (which will not be explored further here).

Now, consider two observations *i* and *i'* that differ in their covariate values, resulting in linear predictors:

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}$$

and:

$$\eta_{i'} = \beta_1 x_{i'1} + \beta_2 x_{i'2} + \ldots + \beta_k x_{i'k}$$

The hazard ratio for these two observations is therefore:

$$h_i(t) / h_{i'}(t)$$

$$=$$

$$h_0(t)e^{\eta i} / h_0(t)e^{\eta i'}$$

$$=$$

$$e^{\eta i} / e^{\eta i'}$$

hence the *proportional-hazards* moniker.

The most important points in the forgoing account of the proportional-hazards model are: a) that the baseline hazards must be proportional; b) that covariates must have a log-linear relationship to the hazard; and c) that because the Cox regression model is log-linear, the hazard ratio can be directly interpreted as a multiplicative relative risk.

5.2 Proportional-hazards modelling of the study data

Based on the results presented in Chapter 4, any considered statistical model of the tubal sterilisation failure data in this study needed to include terms for age, public/private status, tertile of hospital sterilisation workload and financial year of sterilisation. It was tempting to enter a dummy variable for individual hospitals, and possibly an interaction term for individual hospital and year, in order to account for the apparent "special variation" noted in Section 4.5.4. This was attempted, but the

model failed to converge, probably due to the very sparse nature of the failures when cross-classified in this manner. Failure of convergence with very sparse data is a known problem with the partial likelihood fitting process used in Cox regression, a process which is intrinsically less efficient than the maximum-likelihood estimation method used to fit other types of generalised linear models. More advanced statistical models may be able to be fitted to such sparse data, but the use of these is beyond the scope of this study. Therefore, a simpler model using the covariates listed above, without interaction terms, was initially fitted, giving the results shown below. The `coxph` function in the `survival` library for R was used, and its raw output is shown.

```
Call:
coxph(formula = Surv(followup.weeks, censored == 0) ~ hosp.tertile +
    insgrp.at.ster + age.at.ster + ster.fyear, data = falls)

  n=65448 (51 observations deleted due to missingness)
                          coef exp(coef) se(coef)        z       p
hosp.tertile2nd       -0.0502     0.951  0.08448  -0.5940 0.55000
hosp.tertile3rd       -0.2613     0.770  0.09147  -2.8560 0.00430
insgrp.at.sterPrivate -0.1360     0.873  0.15331  -0.8872 0.37000
insgrp.at.sterPublic  -0.2557     0.774  0.14051  -1.8199 0.06900
age.at.ster           -0.0924     0.912  0.00678 -13.6352 0.00000
ster.fyear1993/1994    0.1649     1.179  0.13023   1.2660 0.21000
ster.fyear1994/1995    0.1831     1.201  0.13396   1.3665 0.17000
ster.fyear1995/1996    0.2746     1.316  0.13377   2.0529 0.04000
ster.fyear1996/1997    0.4028     1.496  0.13655   2.9498 0.00320
ster.fyear1997/1998    0.5146     1.673  0.14062   3.6598 0.00025
ster.fyear1998/1999    0.3392     1.404  0.15884   2.1356 0.03300
ster.fyear1999/2000   -0.0028     0.997  0.25191  -0.0111 0.99000

                      exp(coef) exp(-coef) lower .95 upper .95
hosp.tertile2nd           0.951      1.051     0.806     1.122
hosp.tertile3rd           0.770      1.299     0.644     0.921
insgrp.at.sterPrivate     0.873      1.146     0.646     1.179
insgrp.at.sterPublic      0.774      1.291     0.588     1.020
age.at.ster               0.912      1.097     0.900     0.924
ster.fyear1993/1994       1.179      0.848     0.914     1.522
ster.fyear1994/1995       1.201      0.833     0.924     1.561
ster.fyear1995/1996       1.316      0.760     1.013     1.711
ster.fyear1996/1997       1.496      0.668     1.145     1.955
ster.fyear1997/1998       1.673      0.598     1.270     2.204
ster.fyear1998/1999       1.404      0.712     1.028     1.917
ster.fyear1999/2000       0.997      1.003     0.609     1.634

Rsquare= 0.003   (max possible= 0.231 )
Likelihood ratio test= 219  on 12 df,   p=0
Wald test            = 214  on 12 df,   p=0
Score (logrank) test = 216  on 12 df,   p=0
```

The likelihood ratio, Wald, and log-rank (score) chi square statistics at the bottom of the output are equivalent tests of the overall null hypothesis that all of the regression coefficients are zero. In this instance, the test statistics are in close agreement, and the null hypothesis is clearly rejected. The R^2 value is a measure of goodness-of-fit: not particularly good in this case. In the upper panel of results, the column labelled `z` is an asymptotically normal Wald statistic for each parameter estimate, with the corresponding p-value shown to its right. The β estimates for the 3rd tertile of hospital sterilisation workload, for age and for the years 1996/97 and 1997/98 are all significant at the 0.05 level, while the p-value for public patient status was not quite significant (p=0.07). These results correspond with those observed in Chapter 4, and confirm that even after adjustment for confounding, these covariates continue to exert a statistically significant (or almost significant) influence on the rate of sterilisation failure.

It was also hypothesised in Chapter 4 that there may be effect modification as well as confounding between these covariates. To investigate this, interaction terms between each pair of covariates were introduced into the model, and the overall model fit and the parameter estimates for each of these interaction terms were examined and compared to those of a fully-saturated model. None of the interaction terms improved the model. Therefore the initial, parsimonious, first-order model was retained.

5.3 Checking the assumptions of the proportional-hazards model

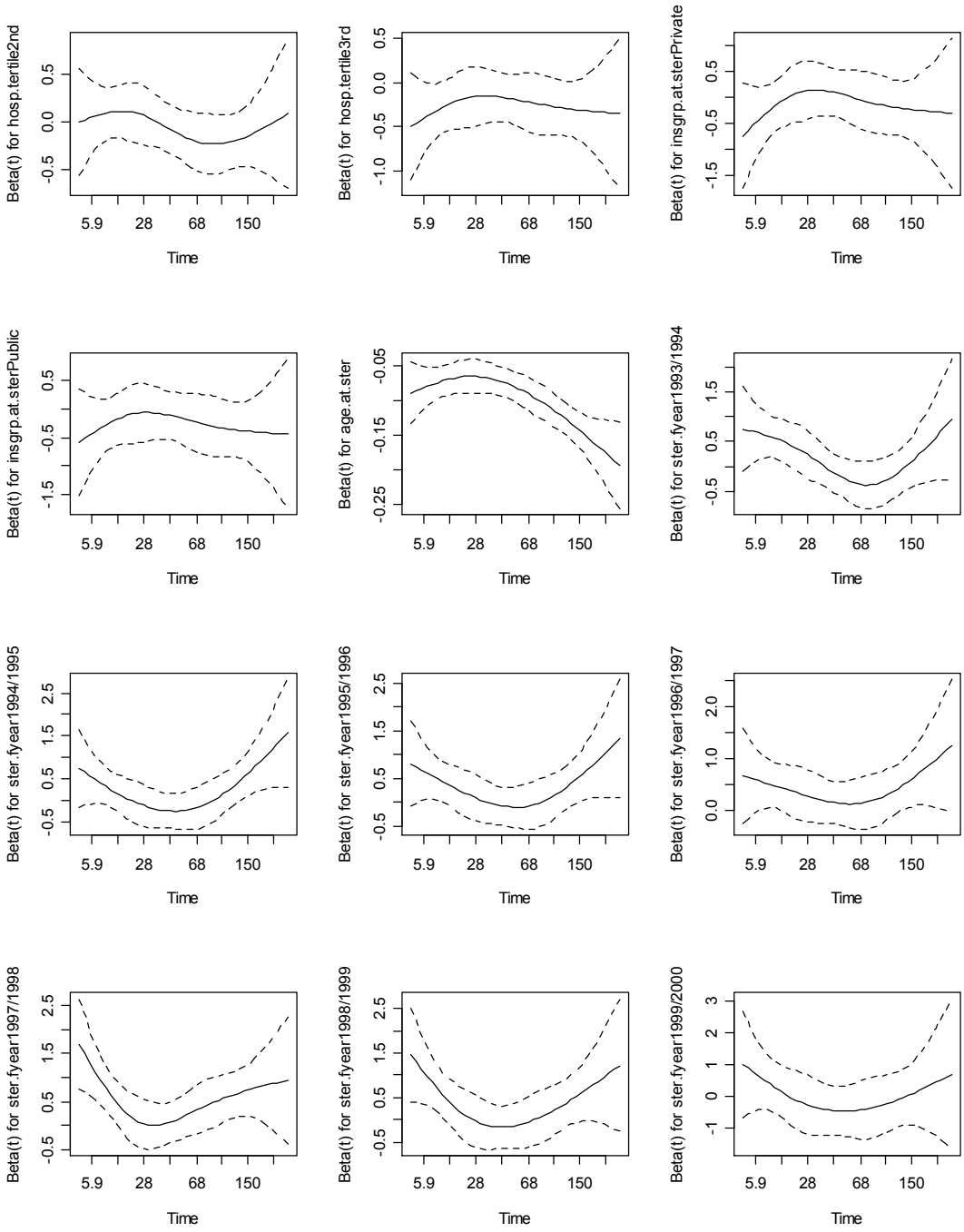
Tests and graphical diagnostics for the proportional-hazards assumption are typically based on *scaled Schoenfeld residuals*. Discussion of the theoretical basis of these tests is beyond the scope of this work. However, the `cox.zph` function in the `survival` library conveniently calculates tests of the proportional-hazards

assumption for each covariate, by correlating the corresponding set of scaled Schoenfeld residuals with time, transformed using the survival function derived from the Kaplan-Meier estimator:

```
> cox.zph(cox.fall)
                           rho    chisq        p
hosp.tertile2nd       -0.03757  1.10316 0.293573
hosp.tertile3rd        0.00414  0.01357 0.907267
insgrp.at.sterPrivate  0.00795  0.04852 0.825650
insgrp.at.sterPublic  -0.00601  0.02815 0.866746
age.at.ster           -0.13417 14.10318 0.000173
ster.fyear1993/1994   -0.05982  2.80620 0.093901
ster.fyear1994/1995    0.02191  0.38008 0.537560
ster.fyear1995/1996   -0.00221  0.00389 0.950291
ster.fyear1996/1997    0.00609  0.02926 0.864173
ster.fyear1997/1998   -0.01743  0.24231 0.622543
ster.fyear1998/1999   -0.02086  0.34640 0.556159
ster.fyear1999/2000   -0.02276  0.40441 0.524818
GLOBAL                     NA 22.64790 0.030868
```

These results indicated that there was strong evidence of non-proportionality of hazard for the age variable, which was also pushing the global test into statistical significance. Smoothed plots of the scaled Schoenfeld residuals against transformed time (using Kaplan-Meier K(t)) for each covariate are shown in Figure 15 below. The usual display of individual residuals was suppressed in these plots because the large number of individual residual points would have otherwise obscured the lines for the smoothed residuals. These plots clearly illustrate the significant departure from proportionality for age as well as less severe departures for the other covariates. Ideally the smoothed plot line should be straight and horizontal.

Figure 15 - Smoothed plots of the scaled Schoenfeld residuals against transformed time (using Kaplan-Meier K(t))for each covariate in the final model



One method for dealing with such departures is to enter interaction terms between time and the covariates with non-proportional hazards into the model. An alternative method involves fitting many models, stratified by a discrete (categorical) transformation of the offending covariates, and then deriving pooled parameter estimates from the ensemble of resulting models. This was tried, with the following results:

```
Call:

coxph(formula = Surv(followup.weeks, censored == 0) ~ hosp.tertile +
    insgrp.at.ster + strata(agegrp.at.ster) + ster.fyear, data =
falls)

  n=65448 (51 observations deleted due to missingness)
                          coef exp(coef) se(coef)       z      p
hosp.tertile2nd       -0.0326     0.968   0.0844 -0.3864 0.7000
hosp.tertile3rd       -0.2470     0.781   0.0915 -2.6995 0.0069
insgrp.at.sterPrivate -0.1979     0.820   0.1531 -1.2926 0.2000
insgrp.at.sterPublic  -0.1575     0.854   0.1404 -1.1219 0.2600
ster.fyear1993/1994    0.1567     1.170   0.1303  1.2026 0.2300
ster.fyear1994/1995    0.1550     1.168   0.1340  1.1568 0.2500
ster.fyear1995/1996    0.2312     1.260   0.1338  1.7280 0.0840
ster.fyear1996/1997    0.3450     1.412   0.1365  2.5270 0.0120
ster.fyear1997/1998    0.4405     1.553   0.1405  3.1342 0.0017
ster.fyear1998/1999    0.2604     1.297   0.1588  1.6397 0.1000
ster.fyear1999/2000   -0.0208     0.979   0.2517 -0.0825 0.9300

                      exp(coef) exp(-coef) lower .95 upper .95
hosp.tertile2nd           0.968      1.033     0.820     1.142
hosp.tertile3rd           0.781      1.280     0.653     0.935
insgrp.at.sterPrivate     0.820      1.219     0.608     1.108
insgrp.at.sterPublic      0.854      1.171     0.649     1.125
ster.fyear1993/1994       1.170      0.855     0.906     1.510
ster.fyear1994/1995       1.168      0.856     0.898     1.518
ster.fyear1995/1996       1.260      0.794     0.969     1.638
ster.fyear1996/1997       1.412      0.708     1.081     1.845
ster.fyear1997/1998       1.553      0.644     1.179     2.046
ster.fyear1998/1999       1.297      0.771     0.950     1.771
ster.fyear1999/2000       0.979      1.021     0.598     1.604

Rsquare= 0   (max possible= 0.202 )
Likelihood ratio test= 24.8  on 11 df,   p=0.0099
Wald test            = 24.4  on 11 df,   p=0.0110
Score (logrank) test = 24.6  on 11 df,   p=0.0104

> cox.zph(cox.falls)
                            rho    chisq      p
hosp.tertile2nd       -0.032626 0.832740 0.3615
hosp.tertile3rd        0.007772 0.047812 0.8269
insgrp.at.sterPrivate -0.002443 0.004581 0.9460
insgrp.at.sterPublic   0.012110 0.114016 0.7356
ster.fyear1993/1994   -0.059781 2.796784 0.0945
ster.fyear1994/1995    0.018936 0.284131 0.5940
ster.fyear1995/1996   -0.006236 0.030947 0.8604
ster.fyear1996/1997    0.000884 0.000616 0.9802
ster.fyear1997/1998   -0.025089 0.501687 0.4788
ster.fyear1998/1999   -0.027617 0.608099 0.4355
ster.fyear1999/2000   -0.020649 0.332277 0.5643
GLOBAL                      NA 8.747158 0.6452
```

These results indicate that such stratification by five-year age groups removes the most egregious source of departure from non-proportionality, albeit at the expense of the loss of regression parameter estimates for the effect of age. Comparison of the significance of and parameter estimates for the remaining covariate values with those

from the initial model revealed only minor changes which did not appreciably affect any conclusions which could be drawn from the models.

Another potential problem is lack of linearity in the relationship between covariates and the hazard. A plot of Martingale residuals against the age covariate (the only continuous covariate in the model) was used to assess this: it revealed substantial non-linearity. The usual approach to this problem is to either categorise the continuous covariate, which loses information, or to try various logarithmic and/or power transformations of the covariate. Recently the use of fractional polynomials, which is a class of polynomial transformations which include negative and fractional powers, has become popular.[195] The `mfp` library for R provides automated backwards elimination, starting with the most complex fractional polynomial in the family of fractional polynomials specified for particular covariates, and uses tests of overall model fit and test for linearity of that covariate in order to find the simplest fractional polynomial transformation which still provides an adequate fit and which satisfies linearity requirements. This was used with the age covariate, with the m=2 family of fractions polynomials specified. The best transformation was found to be the (1,1) fractional polynomial, which has the form x + x*ln*x,. This transformation is also known as the Box-Tidwell transformation. By using it on the age covariate, the following model parameters and proportional-hazards diagnostics were obtained:

```
Call:
mfp(formula = Surv(followup.weeks, censored == 0) ~ fp(age.at.ster,
    df = 4, select = 0.05) + hosp.tertile + insgrp.at.ster +
    ster.fyear, data = falls, family = "cox")


Fractional polynomials:
                     df.initial select alpha df.final power1 power2
age.at.ster                   4   0.05  0.05        1      1      .
hosp.tertile2nd               1   1.00  0.05        1      1      .
hosp.tertile3rd               1   1.00  0.05        1      1      .
ster.fyear1993/1994           1   1.00  0.05        1      1      .
ster.fyear1994/1995           1   1.00  0.05        1      1      .
ster.fyear1995/1996           1   1.00  0.05        1      1      .
ster.fyear1996/1997           1   1.00  0.05        1      1      .
```

```
ster.fyear1997/1998            1   1.00  0.05        1      1      .
ster.fyear1998/1999            1   1.00  0.05        1      1      .
ster.fyear1999/2000            1   1.00  0.05        1      1      .
insgrp.at.sterPrivate          1   1.00  0.05        1      1      .
insgrp.at.sterPublic           1   1.00  0.05        1      1      .

                          coef exp(coef) se(coef)        z       p
age.at.ster.1          -0.0924     0.912  0.00678 -13.6352 0.00000
hosp.tertile2nd.1      -0.0502     0.951  0.08448  -0.5940 0.55000
hosp.tertile3rd.1      -0.2613     0.770  0.09147  -2.8560 0.00430
ster.fyear1993/1994.1   0.1649     1.179  0.13023   1.2660 0.21000
ster.fyear1994/1995.1   0.1831     1.201  0.13396   1.3665 0.17000
ster.fyear1995/1996.1   0.2746     1.316  0.13377   2.0529 0.04000
ster.fyear1996/1997.1   0.4028     1.496  0.13655   2.9498 0.00320
ster.fyear1997/1998.1   0.5146     1.673  0.14062   3.6598 0.00025
ster.fyear1998/1999.1   0.3392     1.404  0.15884   2.1356 0.03300
ster.fyear1999/2000.1  -0.0028     0.997  0.25191  -0.0111 0.99000
insgrp.at.sterPrivate.1 -0.1360    0.873  0.15331  -0.8872 0.37000
insgrp.at.sterPublic.1  -0.2557    0.774  0.14051  -1.8199 0.06900

                       exp(coef) exp(-coef) lower .95 upper .95
age.at.ster.1              0.912      1.097     0.900     0.924
hosp.tertile2nd.1          0.951      1.051     0.806     1.122
hosp.tertile3rd.1          0.770      1.299     0.644     0.921
ster.fyear1993/1994.1      1.179      0.848     0.914     1.522
ster.fyear1994/1995.1      1.201      0.833     0.924     1.561
ster.fyear1995/1996.1      1.316      0.760     1.013     1.711
ster.fyear1996/1997.1      1.496      0.668     1.145     1.955
ster.fyear1997/1998.1      1.673      0.598     1.270     2.204
ster.fyear1998/1999.1      1.404      0.712     1.028     1.917
ster.fyear1999/2000.1      0.997      1.003     0.609     1.634
insgrp.at.sterPrivate.1    0.873      1.146     0.646     1.179
insgrp.at.sterPublic.1     0.774      1.291     0.588     1.020

Rsquare= 0.003   (max possible= 0.231 )
Likelihood ratio test= 219  on 12 df,   p=0
Wald test            = 214  on 12 df,   p=0
Score (logrank) test = 216  on 12 df,   p=0
> cox.zph(f)
                            rho    chisq      p
age.at.ster.1          -7.59e-04 2.86e-05 0.9957
hosp.tertile2nd.1      -3.39e-02 8.96e-01 0.3438
hosp.tertile3rd.1       7.44e-03 4.40e-02 0.8339
ster.fyear1993/1994.1  -5.96e-02 2.79e+00 0.0951
ster.fyear1994/1995.1   1.97e-02 3.07e-01 0.5793
ster.fyear1995/1996.1  -5.52e-03 2.42e-02 0.8763
ster.fyear1996/1997.1   1.56e-03 1.94e-03 0.9649
ster.fyear1997/1998.1  -2.36e-02 4.47e-01 0.5038
ster.fyear1998/1999.1  -2.68e-02 5.72e-01 0.4495
ster.fyear1999/2000.1  -2.22e-02 3.84e-01 0.5356
insgrp.at.sterPrivate.1 1.49e-05 1.71e-07 0.9997
insgrp.at.sterPublic.1  6.04e-03 2.84e-02 0.8662
GLOBAL                        NA 8.54e+00 0.7412
```

These results indicate that little evidence for non-proportionality remains after application of the transformation to age. Plots of Martingale and partial residuals for the transformed age covariate also demonstrated good linearity. The fact that the test for non-proportionality is no longer significant after changes to the model

specification for non-linearity in age suggests that this final model is the best overall model, given the available covariates, and that apparent lack of proportionality in the earlier model may have been principally the result of model misspecification.

5.4 Discussion

The results of the final model can be summarised as follows (HR=hazard ratio, 95% confidence intervals shown in brackets):

- The adjusted HR for women who were sterilised in the lowest workload tertile of hospitals (that is, in the least busy) was 0.77 [0.64-0.92] – that is, they experienced, on average, only three-quarters of the risk of sterilisation failure compared to women who were operated on in the busiest third of hospitals.
- Women sterilised in 1996/97 and 1997/98 had adjusted risks of failure which were about 50 percent and 67 percent higher, on average, than women sterilised in 1992/93 (HR=1.50 [1.15-1.96] and 1.67 [1.27-2.20] for 1996/97 and 1997/98 respectively).
- Increasing age at sterilisation significantly reduced the probability of sterilisation failure. Results from the unadjusted Kaplan-Meier analysis suggested that the risk ratio for age was of the same order of magnitude as that found in the US CREST study.
- When stratification by age group and adjustment for hospital workload tertile was used, the private/public patient status covariate was no longer a statistically significant predictor of failure risk, which added weight to the hypothesis that the differences observed between private and public patient

groups was substantially due to differences in age and/or the smaller size of the hospitals in which private patients tended to be treated.

This Chapter has also illustrated some of the difficulties in fitting a commonly used form of multivariate statistical model to real-life linked data containing only sparse outcomes.

CHAPTER 6

Concluding remarks

This Chapter provides a summary of the methodical issues encountered in undertaking this study and a brief overview of its findings in the context of other studies of tubal sterilisation failure. The potential social, psychological and health service implications of the study results are briefly considered, and some recommendations for further studies and for mechanisms for monitoring the incidence of surgical sterilisation failure are made.

6.1 Findings of the literature review

The findings of the literature reviews have already been summarised in Section 2.10 and will not be repeated at length here. It suffices to say that many studies of failure of tubal sterilisation appeared in the literature through until the late 1980s, but the majority of these studies were case-series or clinic-based, and thus of quite limited generalisability, and almost all failed to systematically follow-up patients and/or to account for differing lengths of time-under-observation in their analyses. Surprisingly, there has been only one large, well-designed study in which follow-up for more than five years was carried out: the US CREST study. None of the published studies, including the CREST study, were population-based: all relied on attendance at participating clinics or hospitals for subject recruitment, with corresponding potential for selection bias.

6.2 Comments on the study design and record linkage process

Although this study, as record linkage studies go, is of fairly simple design, involving only reflexive ("internal") linkage of data derived from a single administrative data collection, there were a surprisingly large number of factors to be considered with respect to its design and execution.

Careful selection of outcome and censoring end points was necessary: in particular, empirical investigation of the data was needed to establish which ICD or other codes should be used to select event records from the source data sets, and careful consideration of how best to infer transitions of patient state or status from these events was required. In this study, the fact and timing of tubal sterilisation procedures was relatively easy to establish, although the level of available detail about those procedures was limited. However, reliable and principled inference of conception subsequent to those tubal sterilisation procedures was more problematic. Pregnancy, abortion and even delivery of infants were recorded in a multiplicity of ways in the source administrative data set. Although the fact of conception could be inferred from these, it was not possible to estimate the date of conception, especially for pregnancies which ended with induced or spontaneous abortion (or for ectopic pregnancies).

Therefore, there was a theoretical possibility that events were included in the data set that actually related to conceptions that occurred prior to the tubal sterilisation procedure, at least for those events occurring up to several months after the procedure. This is particularly important because the hazard of failure is highest in the first few months after the sterilisation surgery.

Exclusion of conception events occurring in the first 28 days after sterilisation was used to mitigate the possibility of undetected luteal phase pregnancies being counted as failures. However, beta-hCG testing on the day before sterilisation has been routinely used in NSW since the late 1980s, and such pre-surgery testing should detect almost all existing pregnancies and result in cancellation of the procedure.

Therefore, it is unlikely that a significant number of the early failures found in this study were in fact related to pre-existing pregnancy.

Repeated references have been made to the fact that the current study could only estimate a lower bound for the cumulative incidence of tubal sterilisation failure. The use of record linkage for follow-up of patients may result in missed links for both outcome and censoring events, both of which will tend to depress absolute estimates of failure incidence, or it may result in false links, which tend to inflate absolute estimates. Due to the limited geographical extent of the data used (NSW only) and the limited number of matching variables available (in particular the absence of name), it seems probable that missed links outnumbered false links in this study. With this in mind, the probabilistic linkage process was undertaken in a way that would minimise false links, abetted by manual review and further pruning of linked outcome records, post-linkage. For these reasons, it seems very unlikely that the estimates of the absolute incidence of failure obtained in this study are biased in an upwards direction – they are far more likely to be biased downwards and can thus be considered lower bounds for the true values.

It was remarked in Chapter 4 that censoring for reasons other than end-of-observation resulted in only a modest 4.1 per cent reduction in the total person-years in the incidence denominator. The implication was that the additional effort in selecting and linking censoring events in this study may not have been justified. However, the marginal effort in linking censoring events, over and above the effort of linking outcome events, was quite small, and the study design would have been considerably weaker had censoring events not been included.

A related observation is that the overall process of probabilistic record linkage is somewhat technical, rather time-consuming and particularly tedious. Approximately 400 hours were spent merely linking and clerically reviewing the data for this study. The establishment of centralised health record linkage facilities, such as the Health Data Linkage Unit in Western Australia and the recently-established Centre for Health Record Linkage in NSW represent enormous advances, because they allow a wider range of source data to be linked once and then used and re-used for many studies. Apart from avoiding duplication of effort, such centralised linkage ensures that sufficient expertise and best practices are brought to bear on the sometimes tricky technical task of probabilistic record linkage. Such centralisation also allows the Boruch and Cecil "separation" model, to be fully implemented, thus minimising the invasion of privacy which is unavoidably associated with all record linkage studies, perhaps to negligible levels.

Despite the real and theoretical problems canvassed above, the use of record linkage of large, population-based data sets does offer many advantages. It would have been impossible to undertake the current study by any other means. The cost of a large, traditional, prospective cohort study such as the US CREST study is likely to be in the millions of dollars and such studies take well over a decade to complete. Given the declining popularity of tubal sterilisation, it is unlikely that funding for such a study could now be obtained. Nevertheless, tubal sterilisation is still a widely-used form of contraception, and it is important to be able to assess the real-life risk of failure and to monitor the corresponding quality of care in health services undertaking the procedure. Additionally, the use of linked, routinely-collected data sets allows studies such as this one to be population-based. This minimises selection bias and allows the results to be more readily generalised. It also means that much

larger studies can be performed, leading to greater statistical precision and power. As already noted, the current study is approximately twice as large as the largest published study, and over six times larger than the CREST study, which is generally regarded as the definitive study on the subject of failure of tubal sterilisation.

6.3 Findings of this study

Despite the design concerns discussed above, the results of this study appear to be sound. In particular, the magnitudes of the crude and conditional cumulative incidence of failure correspond very well with those reported in the large, high-quality published studies, as illustrated in Figure 10 on page 141. In addition, the shapes of the survival curves and hazard functions were as expected and were free from apparent artefact. The relationship of key covariates to the rate of failure also corresponds to the relationships reported in the CREST study.

Previous studies notwithstanding, the findings of an unconditional cumulative incidence of 1.22 failures per 100 women, with up to eight years of follow-up, and a conditional cumulative incidence of failure of 1.51 failures per 100 women at eight years of follow-up (95% CI 1.39-1.62) are notable in themselves. Manufacturers of Filshie clip devices typically quote incidence of failure of under 0.2 per 100 women in their advertising literature. These rates are derived from unpublished clinical trials conducted for device registration purposes. It is likely that obstetricians and gynaecologists quote such figures to candidates for tubal sterilisation. The reality may be that failure is an order of magnitude more frequent. Even an incidence of 1.5 per cent is quite rare. A busy surgeon is unlikely to perform more than one or two hundred tubal sterilisation procedures each year. This, combined with the fact that not all failures occur immediately after the procedure, may mean that many clinicians

remain unaware that their actual failure rate is considerably higher than that quoted by equipment manufacturers.

Tubal sterilisation is not without risks, because it involves admission to hospital, general anaesthesia and the possibility of short-term complications such as gas embolism during inflation of the abdomen for laparoscopy, or deep venous thrombosis and pulmonary embolism. More importantly, this study found that 14 percent of failures presented as ectopic pregnancies. Ectopic pregnancy is potentially life-threatening and continues to be associated with considerable morbidity and occasional mortality. All these risks, although low in absolute terms, must be weighed against the competing risks and reliability of other forms of contraception, as well as considerations of the amenity of a permanent non-barrier form of contraception. In countries with a high prevalence of HIV infection, the dangers posed by non-barrier contraception and "unsafe" sex must also be considered. Weighing of these risks and benefits is beyond the scope of this study, but such analysis must be predicated on accurate estimates of reliability and failure rates for each method of contraception. This study, or others like it, may provide better estimates of failure rates under contemporary 'real-life' conditions than are currently available.

This study also found that 40 per cent of post-sterilisation pregnancies terminated in abortion, the majority of which were likely to have been induced. Any abortion, but particularly an induced one, is a stressful and traumatic event which may have long-term mental health and social consequences. Similarly, the raising and support of an

unplanned child may, for some people, represent a large, ongoing mental, social or economic burden.

Several predictors of failure were discovered in this study. As other investigators have found, the risk of sterilisation failure declined with age, probably due to changes in underlying fertility. More surprisingly, a lower risk of failure was found in patients who had their surgery in hospitals with a lower workload of tubal sterilisation procedures. A lower hazard was also found for privately insured patients, who tend to be treated more often in private hospitals. These in turn tend to be smaller than public-sector teaching and regional hospitals. Additionally, patients with private insurance were slightly older on average. Thus, three-way confounding was possible. However, multivariate analysis using Cox regression revealed statistically significant decreased hazard ratios for surgery in lower-workload (smaller) hospitals, even after adjusting for age and health insurance status.

The reason for this finding, which is slightly counter-intuitive, may be related to the fact that larger hospitals are also teaching institutions, and thus a significant proportion of tubal sterilisations in them may be performed by relatively inexperienced trainee surgeons. Adequate supervision of such predominantly laparoscopic surgery can be technically challenging.

Also of considerable interest was the finding of an excess of failures in two hospitals, and the fact that these excesses were confined to relatively short periods of one or two years in each hospital. This strongly suggested a transient “special cause” for at least some of these failures, although it is impossible to say from this study whether such a special cause might be equipment- or personnel-related. Risk of failure was

also found to be have increased generally in 1996/97 and 1997/98. This suggested some systematic cause although this study provided no clues as to what that might be.

6.4 Summary and recommendations for further studies

This study has, despite its acknowledged limitations, clearly demonstrated the feasibility of using linked administrative health data to examine relatively rare outcomes of a common surgical procedure on a short-, medium- and long-term basis. Its findings are entirely consistent with those of several large conventional studies.

Furthermore, it is, the to author's knowledge, the only study of tubal sterilisation failures which has used record linkage methods for follow-up, and the only such study which is truly population-based. It is also considerably larger than extant studies.

Given these benefits, it seems sensible that the study be repeated in the near future using linked data from the NSW Centre for Health Record Linkage and/or the Western Australian Health Data Linkage Unit. Both of these facilities have access to, and make use of, name information in their probabilistic linkage processes, resulting in higher quality data sets with fewer missed links and fewer false links. In addition, data from sources other than admitted patient data collections are used: in particular, data from maternal and neonatal data collections and death certificates are also linked. This should provide more complete ascertainment of failure outcomes and improved censoring. In the longer term, it may be possible to use linked data with a national scope, created by a consortium of State and Territory-based health record linkage facilities.

It may also be possible to use data from these ongoing record linkage facilities to continuously monitor failure rates at a jurisdictional or institutional level. Because survivorship at various times after surgery can be estimated from historical data, it is not necessary to wait many years before assessing the performance of a hospital or clinic – the expected number of failures at intervals of months or years after surgery can easily be calculated and compared with the observed number of failures.

In these respects, this study typifies and perhaps serves as a model for a whole class of health services outcome studies which can be performed using linked administrative data.

Finally, it is hoped that this study helps to better inform women considering tubal sterilisation as a contraceptive option.

APPENDICES

Appendix A: The Fellegi-Sunter model of probabilistic record linkage, as extended by Winkler and Jaro.

Consider two populations (sets) of records, **A** and **B**, whose elements are denoted by a and b. We assume that at least some elements are common to **A** and **B**.

The ordered Cartesian product (cross-product) of these two sets :

$$\mathbf{A}\times\mathbf{B} = \{(a,b): a\varepsilon\mathbf{A}, b\varepsilon\mathbf{B}\}$$

must be the union of two disjoint sets of *matches*

$$M = \{(a,b): a=b, a\varepsilon\mathbf{A}, b\varepsilon\mathbf{B}\}$$

and *nonmatches*

$$U = \{(a,b): a\neq b, a\varepsilon\mathbf{A}, b\varepsilon\mathbf{B}\}$$

Values for each data element, a or b, of each member of **A** and **B** are denoted $\alpha(a)$ and $\beta(b)$. A *comparison vector* γ associated with these data value vectors is defined by

$$\gamma[\alpha(a), \beta(b)] \equiv \{ \gamma^1[\alpha(a), \beta(b)], \gamma^2[\alpha(a), \beta(b)], \ldots, \gamma^K[\alpha(a), \beta(b)] \}$$

Thus, each γ^i, i=1,…,K, represents the result of a specific comparison of particular data elements – for example, γ^1 might represent agreement or disagreement on sex, γ^2 might represent the comparison of two surnames, and so on. The function γ on $\mathbf{A}\times\mathbf{B}$ will be denoted by $\gamma(\alpha, \beta)$ or $\gamma(a, b)$ or just γ. The set of all possible realisations of γ for $\mathbf{A}\times\mathbf{B}$ is denoted by Γ.

The conditional probability of $\gamma(a, b)$ if $(a,b)\varepsilon M$ is given by

$$m(\gamma) \equiv \Pr\{ \gamma[\alpha(a), \beta(b)] \mid (a,b)\varepsilon M \}$$

$$= \sum_{(a,b)\varepsilon M} \Pr\{ \gamma[\alpha(a), \beta(b)] \} \cdot \Pr[(a,b) \mid M]$$

Similarly the conditional probability of γ if $(a,b)\varepsilon U$ is denoted $u(\gamma)$.

We process our data and observe a vector of information $\gamma(a,b)$ associated with the pair of records (a,b). We wish to designate the pair as a match (link) and thus an element of set A_1, a possible match (in set A_2) or a non-match (in set A_3). L denotes a linkage rule (a set of probability cut-offs or thresholds) which divides Γ into A_1, A_2 and A_3.

By convention, a *Type 1* error occurs if rule L places $m\varepsilon M$ in A_3, with probability:

$$\Pr(A_3|M) = \sum_{\gamma\in\Gamma} m(\gamma) \cdot P(A_3 \mid \gamma)$$

and a *Type II* error occurs if L places $u\varepsilon U$ in A_1, with probability:

$$\Pr(A_1|U) = \sum_{\gamma\in\Gamma} u(\gamma) \cdot P(A_1 \mid \gamma)$$

Fellegi and Sunter proposed a linkage rule L_O with associated sets A_1, A_2 and A_3. which is optimal in the sense that:

> Theorem (Fellegi-Sunter 1969). Let L' be a linkage rule for determining associated sets A_1', A_2' and A_3' such that it has the same probabilities of *Type I* and *Type II* errors as decision rule L_O: $\Pr(A_3' \mid M) = \Pr(A_3 \mid M)$ and $\Pr(A_1' \mid U) = \Pr(A_1 \mid U)$. Then L_O is optimal in that $\Pr(A_2 \mid M) \leq \Pr(A_2' \mid M)$ and $\Pr(A_2 \mid U) \leq \Pr(A_2' \mid U)$.

To describe L_O we compute the following likelihood ratio:

$$R \equiv \quad R[\gamma(a, b)] = m(\gamma) / u(\gamma)$$

$$= \quad \frac{Pr\{ \gamma[\alpha(a), \beta(b)] \mid (a,b)\varepsilon M \}}{Pr\{ \gamma[\alpha(a), \beta(b)] \mid (a,b)\varepsilon U \}}$$

If γ represents a comparison of K data elements or fields, then there are 2^K probabilities of the form $m(\gamma)$. If the comparison vector γ represents agreements on the K fields, then we would expect such agreements to occur more often for the set of matches M than for the set of non-matches U – thus the ratio R would be large. Conversely, if γ consists of disagreements, R would be small.

The Fellegi-Sunter linkage rule L_O takes the form:

If $R > T_\mu$ then consider (a,b) as a match and place it in set A_1

If $T_\lambda \leq R \leq T_\mu$ then consider (a,b) as a possible match and place it in set A_2

If $R < T_\lambda$ then consider (a,b) as a non-match and place in set A_3

The cut-offs T_λ and T_μ are determined by the desired error bounds μ and λ on the false match rates (incorrect links) and the false nonmatch rates (missed links) respectively.

It is possible to compute the critical values T_μ and T_μ for the linkage rule L_O from the data itself, if conditional independence is assumed:

$$m(\gamma) = m_1(\gamma^1) \cdot m_2(\gamma^2) \cdot \ldots \cdot m_K(\gamma^K)$$

and

$$u(\gamma) = u_1(\gamma^1) \cdot u_2(\gamma^2) \cdot \ldots \cdot u_K(\gamma^K)$$

where for i = 1,2,…,K

$$m_i(\gamma^i) = Pr\{\gamma^i \mid (a,b)\varepsilon M\}$$

and

$$u_i(\gamma^i) = Pr\{\gamma^i \mid (a,b)\varepsilon U\}$$

and secondly by using $\log_2$ to compute the ratio:

$$W \equiv W(\gamma) = \log_2[m(\gamma) / u(\gamma)] = W^1 + W^2 + \ldots + W^K$$

where $W^i = \log_2[m_i(\gamma^i) / u_i(\gamma^i)]$ for i = 1,2, …, K.

W is known as the *total comparison weight* associated with a pair of records, and W_i for i = 1,2,…,K are known as the *individual comparison weights*.

Fellegi and Sunter[135] demonstrated algebraic solutions for determining the optimal rule L_O for the case in which simple agree/disagree comparisons were being made for only three fields (data items). Winkler[136] and Jaro[138] introduced the use of expectation maximisation (EM) to find optimal numerical solutions for (the usual) situation in which more than three data items were to be compared.

Fellegi and Sunter also described how this model can be extended to use value frequency information as follows. Let the true frequencies of a given value (say a given name) in sets **A** and **B**, respectively be:

$$f_1, f_2,\ldots,f_m; \sum_{j=1}^{m} f_j = N_A$$

and

$$g_1, g_2,\ldots,g_m; \sum_{j=1}^{m} g_j = N_B$$

Let the corresponding true frequencies of values in **A** ∩ **B** (M) be:

$h_1, h_2, \ldots, h_m; \sum_{j=1}^{m} h_j = N_M$

Note that $h_j \leq \min(f_j, g_j)$ for j=1,2,…,m. For most data sets, it is safe to use

$h_j = \min(f_j, g_j)$ if $f_j > 1$ or $g_j > 1$; or

$h_j = 2/3$ otherwise.

In other words, if only one pair is observed to agree on a specific value of a name, then that pair has 2/3 chance of being a match and 1/3 chance of being a non-match.

The following symbols are also required:

e_A, e_B

The probabilities of a value being incorrectly reported in **A** or **B** respectively (assuming independence with respect to particular values).

e_{A0}, e_B

The probabilities of a missing value in **A** or **B** respectively (independence assumed).

e_T

The probability that a value is differently (but correctly) reported in **A** and **B.**

Thus:

Pr(agrees on *j*th value of data element | M)

$= h_j(1-e_A)(1-e_B)(1-e_T)(1-e_{A0})(1-e_{B0})/N_M$

$= h_j (1- e_A-e_B-e_T-e_{A0}-e_{B0})/N_M$

Pr(value of data element disagrees | M)

$= [1-(1-e_A)(1-e_B)(1-e_T)] (1-e_{A0})(1-e_{B0})/N_M$

$= e_A + e_B + e_T$

Pr(value of data element missing in either **A** or **B** | M)

$= 1-(1-e_{A0})(1-e_{B0})$

$= e_{A0} + e_{B0}$

Pr(agrees on *j*th value of data element | U)

$= (f_j \cdot g_j - h_j)(1-e_A)(1-e_B)(1-e_T)(1-e_{A0})(1-e_{B0})/(N_A \cdot N_B - N_M)$

$= (f_j \cdot g_j - h_j)(1-e_A-e_B-e_T-e_{A0}-e_{B0})/(N_A \cdot N_B - N_M)$

Pr(value of data element disagrees | U)

$$= [1-(1-e_A)(1-e_B)(1-e_T)\sum_{j=1}^{m} (f_j \cdot g_j - h_j) /(N_A \cdot N_B - N_M)](1-e_{A0})(1-e_{B0})$$

$$= [1-(1-e_A-e_B-e_T)\sum_{j=1}^{m} (f_j \cdot g_j - h_j) /(N_A \cdot N_B - N_M)](1-e_{A0})(1-e_{B0})$$

Pr(value of data element missing in either **A** or **B** | U)

$= 1-(1-e_{A0})(1-e_{B0})$

$= e_{A0} + e_{B0}$

The relative frequency scaling weights for agreement on the *j*th specific name value, j=1,2,…,m, are given by:

$wgt(j) = h_j \cdot (N_A \cdot N_B - N_M) / ((f_j \cdot g_j - h_j) \cdot N_M)$ if $f_j > 1$ or $g_j > 1$

or by:

$wgt(j) = 2(N_A \cdot N_B - N_M) / N_M$ if $f_j = 1$ and $g_j = 1$

Note that e_{A0} and e_{B0} can be estimated directly from the **A** and **B** files. However, to estimate e_A, e_B and e_T the intersection of **A** and **B** (that is, the linked records) must be known, and thus they cannot be estimated directly. Typically, initial guesses based on past experience are used. However, use of the EM algorithm allows direct estimation of Pr(value of data element disagrees | M) and therefore approximate estimation of $e_A+e_B+e_T$.

Further adjustment of weights for typographical errors and to account for the use of "blocking" (see Section 3.2.5) are also used but will not be described here.

Appendix B: Excerpt from the *AutoStan* address standardisation patterns used in this study.

The following street address standardisation patterns were loosely based on a set provided to the author by Matthew Jaro, creator of the *AutoStan* software with which they are used. They were extensively modified by the author in the early 1990s, and have been added to and adjusted on occasion since then by Kim Lim of the Centre for Epidemiology and Research, NSW Department of Health. A similar file of patterns was used to standardise locality information to be linked.

AutoStan is a re-entrant tokeniser and parser which uses a slightly idiosyncratic modification of the standard and widely-used Backus-Naur Form (BNF) regular expression syntax to operate on address or names strings as a sequence of tagged tokens and/or as strings.[140] A full explanation of the syntax is beyond the scope of this work, but essentially each block of code contains a regular expression-like matching rule in its first line, and then a series of actions which are performed if that rule fires. Matching may be by type of token. Token types are assigned to words using look-up tables (not shown). The actions may re-write or modify the sequence of tokens for further processing, or output information to a results file. The sequence of tokens is then passed to the next rule (hence its description as "re-entrant"). A small sample of rules and actions is presented here to provide the reader with an insight into the methods used for file preparation in this record linkage study. The complete STREET.PAT file contains 8,396 lines, and the corresponding LOCALITY.Pat file for standardising locality (suburb, town) information contains 3,146 lines.

```
; STREET.PAT - ISC Address Standardiser - street process patterns
; original for US addresses by Matt Jaro, modified by Tim Churches
; for Australian address forms and maintained by Kim Lim.
```

```
; Token Classification is as follows:
; 0 - Null token (ignored)
; A - Adjective (Great, Little etc)
; B - (Mail) Bag or Box
; C - Cardinal numbers
; D - Compass direction
; E - Australian State and Territories
; F - "Office"
; G - "Mail"
; H - Geographical features (River, Mount etc)
; I - Corner, intersection
; J - "and" etc
; K - Locality qualifier (upper, lower, heights etc)
; M - Unit type (Unit, Apartment, Block etc)
; O - Ordinals
; P - "Post", "PO" etc
; Q - "The"
; R - Dummy token used for placeholding
; S - Field separator "|" (upright bar)
; T - Wayfare type
; U - "Unknown", "NA"
; V - "Via", "By"
; W - "Lot"
; X - Institutions (University, Nursing Home etc)
; Y - "Tens" cardinals (Eighty, Ninety etc)
; Z - "Floor" or "Level"
;
; Calculate Soundex and reverse Soundex for Street names,
; sub-address names and localities (main and alternative)
\POST_START
SOUNDEX {WN} {XS}
RSOUNDEX {WN} {XR}
SOUNDEX {SA} {SS}
RSOUNDEX {SA} {SR}
SOUNDEX {LO} {XL}
RSOUNDEX {LO} {RL}
SOUNDEX {AL} {AX}
RSOUNDEX {AL} {AR}
\POST_END

I | ? | J | ? ; Corner xyz and abc sts
COPY [2] temp
CONCAT "/" temp
CONCAT [4] temp
RETYPE [1] 0
RETYPE [2] ? temp
COPY temp {WN}
RETYPE [3] 0
RETYPE [4] 0

I | H | J | ? ; Corner xyz and abc sts
COPY [2] temp
CONCAT "/" temp
CONCAT [4] temp
RETYPE [1] 0
RETYPE [2] ? temp
COPY temp {WN}
RETYPE [3] 0
RETYPE [4] 0

*? | J | ?  ; xyz and abc sts
COPY_S [1] temp
CONCAT "/" temp
CONCAT [3] temp
```

```
RETYPE [1] ? temp
COPY temp {WN}
RETYPE [1] 0
RETYPE [2] 0
RETYPE [3] 0

; Process single letter slash word as abbreviation eg P/WOOD for
PITTWOOD
*+ [{}LEN=1] | / | ?[{}LEN>1] ;Single letter slash word as
abbreviation
RETYPE [1] ?
RETYPE [2] ?
RETYPE [3] ?

; Cope with trailing R, S, C etc.
+ [{}="R"] | $  ;Trailing R
RETYPE [1] T "RD" "ROAD"

+ [{}="S"] | $  ;Trailing S
RETYPE [1] T "ST" "STREET"

+ [{}="C"] | $  ;Trailing C
RETYPE [1] T "CRES" "CRES"
E="C" | $ ;TRAILING C
RETYPE [1] T "CRES" "CRES"

+ [{}="D"] | $  ;Trailing D
RETYPE [1] T "DR" "DRIVE"

+ [{}="A"] | $ ;Trailing A
RETYPE [1] T "AVE" "AVE"

E="A" | $ ;TRAILING A
RETYPE [1] T "AVE" "AVE"

E="T" | $ ;TRAILING T
RETYPE [1] T "TCE" "TCE"

+="L" | $ ;TRAILING L
RETYPE [1] T "LANE" "LANE"

M | ^ | ^ | - | ^ | ? |T; eg UNIT 10 12-14 WHATEVER ST
COPY_A [1] {UT}
COPY [2] {UV}
COPY [3] {HN}
COPY [5] {HR}
COPY_S [6] {WN}
COPY_A [7] {WT}
RETYPE [1] 0
RETYPE [2] 0
RETYPE [3] 0

< | Q | D | T; eg R677 THE NORTHERN RD
COPY [1] {HN}
COPY [2] temp
CONCAT " " temp
CONCAT [3] temp
MOVE temp {WN}
COPY_A [4] {WT}
RETYPE [1] 0
RETYPE [2] 0
RETYPE [3] 0

...
```

Appendix C. ICD-9-CM and ICD-10AM diagnosis and procedure codes used to select ISC records for this study

* denotes a code which was subsequently considered insufficiently specific, and which not used to infer conception in final data set.

Table 18 - ICD-9-CM and ICD-10AM diagnosis and procedure codes used to select ISC records for this study

	ICD-9-CM (1992/93 to 1997/98)		ICD-10-AM (1998/99 to 1999/2000)	
Category	**Codes**	**Description**	**Codes**	**Description**
Diagnosis and procedure codes indicative of tubal sterilisation	V25.2	Sterilisation (Interruption of fallopian tubes)	Z30.2	Sterilisation
	66.2	Bilateral endoscopic destruction or occlusion of fallopian tubes	35688-00	Laparoscopic sterilisation
	66.3	Other bilateral destruction or occlusion of fallopian tubes	35688-01	Sterilisation via vaginal approach
	66.5	Total bilateral salpingectomy	35688-02	Sterilisation via open abdominal approach
	66.63	Bilateral partial salpingectomy	35688-03	Laproscopic electrodestruction of fallopian tubes
	66.64	Electrodestruction of fallopian tube	35688-04	Electrodestruction of fallopian tubes
	66.69	Other bilateral partial salpingectomy	35638-08	Laparoscopic partial bilateral salpingectomy
			35717-02	Partial bilateral salpingectomy
			35638-10	Laparoscopic bilateral salpingectomy
			35717-03	Bilateral salpingectomy

	ICD-9-CM (1992/93 to 1997/98)		ICD-10-AM (1998/99 to 1999/2000)	
Category	**Codes**	**Description**	**Codes**	**Description**
Diagnosis and procedure codes indicative or suggestive of conception (failure events)	181	Malignant neoplasm of placenta	C58	Malignant neoplasm of placenta
	236.1	Malignant hydatidiform mole	D39.2	Malignant hydatidiform mole
	253.2	Postpartum pituitary necrosis	E23.0	Postpartum pituitary necrosis
	630-633	Ectopic and molar pregnancy	O00-O02	Ectopic pregnancy, hydatidiform mole and other abnormal products of conception
	634-639	Other pregnancy with abortive outcome	O03-O08	Spontaneous, medical, unspecified, failed attempted and other abortions
	640-648	Complications related to pregnancy	O10-O16	Oedema, proteinuria and hypertensive disorders in pregnancy, childbirth and the puerperium
			O20-O29	Other maternal disorders predominantly related to pregnancy
			F53	Puerperial mental disturbance
			O09	Duration of pregnancy
	650-659	Normal delivery, and other indications for care in pregnancy, labour and delivery	O30-O48 and O80-O84	Maternal care related to the foetus and amniotic cavity and possible delivery problems; and normal delivery of singleton and multiple births
	660-669	Complications occurring mainly in the course of labour and delivery	O60-O75	Complications of labour and delivery
	670-676	Complications of the puerperium	O85-O92	Complications predominantly related to the puerperium
			A34	Obstetrical tetanus
			O95-O99	Other Obstetric conditions NEC
	V22	Normal pregnancy	Z32.1	Pregnancy (confirmed)
			Z33	Pregnant state NOS

Category	ICD-9-CM (1992/93 to 1997/98)		ICD-10-AM (1998/99 to 1999/2000)	
	Codes	**Description**	**Codes**	**Description**
	V23	Supervision of high-risk pregnancy	Z35	Supervision of a high risk pregnancy
	V24	Postpartum care and examination	Z39	Postpartum care and examination
	V27	Outcome of delivery	Z37	Outcome of delivery
	V28	Antenatal screening	Z36	Antenatal screening
			M83.0	Puerperial osteomalacia
	66.0*, 66.62	Salpingotomy, salpingectomy with removal of tubal pregnancy. Note: 66.0 salpingotomy subsequently considered insufficiently specific, not used to infer conception in final data set)	35677-01, 35674-00, 35674-01, 35677-02, 35677-03, 35677-00, 35677-05	Salpingectomy with removal of tubal pregnancy
	66.90	Foetotoxic management for removal of ectopic pregnancy at any site	35678-00, 35678-01	Foetotoxic management for removal of ectopic pregnancy at any site
	68.0	Hysterotomy with removal of hydatidiform mole		
	69.0*	D&C (Note: subsequently considered insufficiently specific, not used to infer conception in final data set)	35640-00*, 35640-01*	D&C (Note: subsequently considered insufficiently specific, not used to infer conception in final data set)
	69.5*	Aspiration and curettage (Note: subsequently considered insufficiently specific, not used to infer conception in final data set)	35643-00*	Aspiration and curettage (Note: subsequently considered insufficiently specific, not used to infer conception in final data set)
	69.51, 69.52	Aspiration curettage for termination or following delivery/abortion	16564-00, 35643-00,	Aspiration curettage for termination or following delivery/abortion

Category	ICD-9-CM (1992/93 to 1997/98)		ICD-10-AM (1998/99 to 1999/2000)	
	Codes	**Description**	**Codes**	**Description**
			35643-01	
	69.6*	Menstrual extraction or regulation (Note: subsequently considered insufficiently specific, not used to infer conception in final data set)	Z30.3*	Menstrual extraction or regulation (Note: subsequently considered insufficiently specific, not used to infer conception in final data set)
	69.93	Insertion of laminaria		
	72-75	Obstetric procedures	Codes shown at right	Obstetric procedures (90460-00, 90461-00, 90462-00, 16600-00, 16618-00, 16621-00, 16603-00, 16606-00, 16627-00, 90463-00, 16609-00, 16612-00, 16615-00 16624-00, 90464-00, 90486-00, 90486-01, 90486-02, 90465-00, 90465-01, 90465-02, 90465-03, 90465-04, 90465-05, 90466-00, 90466-01, 90466-02, 90467-00, 90468-00, 90468-01, 90468-02, 90468-03, 90468-04, 90468-05, 90469-00, 90469-01, 90470-00, 90470-01, 90470-02, 90470-03, 90470-04, 16520-00, 16520-01, 16520-02, 16520-03, 16514-00, 16514-01, 90471-00, 90471-01, 90471-02, 90471-03, 90471-04, 90471-05, 90471-06, 90472-00, 90473-00, 90474-00, 90475-00, 90476-00, 90477-00, 90478-00, 16571-00, 90479-00, 90480-00, 90480-01, 90485-00, 90481-00, 16573-00, 90482-00, 16564-00, 16564-01, 16570-00, 16570-01, 90483-00, 90484-00, 90484-01, 90484-02, 16567-00)

	ICD-9-CM (1992/93 to 1997/98)		ICD-10-AM (1998/99 to 1999/2000)	
Category	**Codes**	**Description**	**Codes**	**Description**
	96.49	Insertion of prostaglandin suppository for induction of abortion without labour	90461-00, 90462-00, 90465-01	Insertion of prostaglandin suppository for induction of abortion without labour
Codes indicative or suggestive of infertility even if Fallopian tubes are patent (censoring events)	65.5	Bilateral oophorectomy	35717-02, 35638-08, 35638-03, 35638-12, 35717-01, 35638-10, 35717-04, 35717-03, 35717-05	Oophorectomy and salpingectomy
	65.6	Bilateral salpingo-oophorectomy		
	68.3	Subtotal abdominal hysterectomy	35653-00	Subtotal abdominal hysterectomy
	68.4	Total abdominal hysterectomy	35653-01, 35653-02, 35653-03	Total abdominal hysterectomy
	68.5	Vaginal hysterectomy	35657-00, 35750-00, 35756-00, 35673-00, 35673-01, 35753-00, 35753-00,	Vaginal hysterectomy

	ICD-9-CM (1992/93 to 1997/98)		ICD-10-AM (1998/99 to 1999/2000)	
Category	**Codes**	**Description**	**Codes**	**Description**
			35756-01, 35756-02, 35667-01, 35664-01	
	68.6	Radical abdominal hysterectomy	35661-00	Radical abdominal hysterectomy
	68.7	Radical vaginal hysterectomy	35670-00 , 35667-00	Radical vaginal hysterectomy
	68.8	Pelvic evisceration	35664-00	Pelvic evisceration
	68.9	Other and unspecified hysterectomy	90443-00	Other and unspecified hysterectomy
Codes indicative or suggestive of attempts at reversal of tubal sterilisation (censoring events)	V26.0	Tuboplasty after previous sterilisation	Z31.0	Tuboplasty after previous sterilisation
	66.7	Repair of fallopian tube	35697-00	Microsurgical salpingoplasty
	66.8	Insufflation of fallopian tube	35694-00	Laparoscopic salpingoplasty
	66.93	Implantation or replacement of prosthesis of fallopian tube	35694-04	Salpingoplasty
			35694-01	Laparoscopic anastomosis of fallopian tube
			35694-05	Anastomosis of fallopian tube

	ICD-9-CM (1992/93 to 1997/98)		ICD-10-AM (1998/99 to 1999/2000)	
Category	**Codes**	**Description**	**Codes**	**Description**
			35694-03	Laparoscopic salpingostomy
			35694-07	Salpingostomy
			90433-00	Other laparoscopic repair of fallopian tube
			90433-01	Other repair of fallopian tube
			35710-00	Falloposcopy
			35703-01	Therapeutic hydrotubation
			35694-02	Laparoscopic salpingolysis
			35694-06	Salpingolysis
Codes indicative or suggestive of assisted reproduction techniques which bypasses the fallopian tubes (censoring events)	V26.8	Other specified procreative management (IVF, GIFT)	Z31.2	IVF
	69.8	Reproductive medicine procedures	Z31.3	Other assisted fertilisation methods
	69.83	GIFT	13215-00	GIFT
	69.84	Embryo transfer to uterus	13218-00	Embryo transfer to uterus
	69.85	Embryo transfer to fallopian tube	13218-01	Embryo transfer to fallopian tube

Category	ICD-9-CM (1992/93 to 1997/98)		ICD-10-AM (1998/99 to 1999/2000)	
	Codes	Description	Codes	Description
	69.89	Other assisted reproduction NOS (IVF)	13200-00, 13206-00	Assisted reproductive services
			13203-00	Ovulation monitoring services
			13212-00, 13212-01	Oocyte retrieval
			13209-00	Planning and management for assisted reproductive technologies
			13218-02	Other reproductive medicine procedure

Appendix D. *AutoMatch* linkage specifications used for this study

Following is an annotated and abridged version of the *AutoMatch* linkage specification file used for this study. Annotations are italicised.

```
/* "geomatch" specifies that many records in the A file can match with one record in the B file */
PROGRAM GEOMATCH
DICTA endpoints9200  /* file A: conceptions and censoring events */
DICTB sterprocs9200 ; /* file B: tubal sterilisation procedures */

/* Pass 1: compare only records with the same hospital code and MRN in this pass */
BLOCK1  CHAR      Hoscode   hoscode
BLOCK1  CHAR      MRN       MRN
/* fields compared are only hospital code and MRN, with m- and u-probabilities set so the effectively deterministic matching occurs */
MATCH1  CHAR      Hoscode   hoscode  0.99 0.00005
MATCH1  CHAR      MRN       MRN      0.99 0.00000001
;

/*Pass 2: compare only records with the same MRN and the same pseudo-hospital code, which conflates all hospitals in the two Area health Services which had pan-Area MRN systems, this pass */
BLOCK2  CHAR      ahoscode  ahoscode
BLOCK2  CHAR      MRN       MRN
/* add date of birth comparison as an extra check in this pass, as pan-AHS MRN assignment process was somewhat imperfect for several years and some MRNs were re-used for different patients */
MATCH2  CNT_DIFF  Birthd    birthd   0.9  0.0005 1
MATCH2  CHAR      ahoscode  ahoscode 0.9  0.005
MATCH2  char      MRN       MRN      0.99 0.00000001
;

/*Pass 3: compare all records with the same DOB and the same postcode of residence */
BLOCK3  NUMERIC   Birthd    birthd
BLOCK3  CHAR      pc        pc
/* use many different partially-identifying items for the record comparisons */
MATCH3  CNT_DIFF  Birthd    birthd   0.9  0.0005 1 /* Date of birth */
MATCH3  CHAR      Birthpl   birthpl  0.9  0.617 /* Country of birth */
MATCH3  ARRAY CNT_DIFF nums  nums     0.5  0.143 1 /* Compare the Cartesian product of an array of house/unit/street numbers */
MATCH3  CNT_DIFF  hs        hs       0.5  0.444 1 /* Count differences in characters in street number field */
MATCH3  UNCERT    wn        wn       0.9  0.006 700.0 /* Compare Winkler-Jaro similarity co-efficient for street names */
MATCH3  CHAR      wt        wt       0.5  0.276 /* Wayfare (street) type */
MATCH3  CHAR      ut        ut       0.5  0.429 /* Unit/dwelling type *
MATCH3  CHAR      pt        pt       0.5  0.429 /* Postal address comparison */
```

```
MATCH3  ARRAY UNCERT sc      sc       0.9  0.005 700.0 /* Cartesian cross-
comparison of street name fields by string similarity */
MATCH3  UNCERT     st        st       0.5  0.11 700.0
MATCH3  NUMERIC    pc        pc       0.5  0.003 /* Postcode */
MATCH3  ARRAY UNCERT lc      lc       0.5  0.002 700.0 /* Cartesian cross-
comparison of suburb/locality/town fields by string similarity */
;

/*Pass 4: compare all records with the same year and month of birth and the same
phonetic encoding of suburb/locality/town name */
BLOCK4  NUMERIC    birthyear  birthyear
BLOCK4  NUMERIC    birthmonth birthmonth
BLOCK4  CHAR       xl         xl
/* record comparison same as for pass 3 above */
…
;

/*Pass 5: compare all records with the same month of birth and day of month or
birth and same the same phonetic encoding of suburb/locality/town name and same
phonetic encoding of street name*/
BLOCK5  NUMERIC    birthmonth birthmonth
BLOCK5  NUMERIC    birthday   birthday
BLOCK5  CHAR       xl         xl
BLOCK5  CHAR       xs         xs
/* record comparison same as for pass 3 above */
…
;

/*Pass 6: compare all records with the same month of birth and day of month or
birth and same the same phonetic encoding of suburb/locality/town name and same
phonetic encoding of street name*/
BLOCK6  NUMERIC    birthyear  birthyear
BLOCK6  NUMERIC    birthday   birthday
BLOCK6  CHAR       xl         xl
BLOCK6  CHAR       xs         xs
/* record comparison same as for pass 3 above */
…
;

/*Pass 7: compare all records with the same year of birth and same phonetic
encoding of street name in the same Area Health Service*/
BLOCK7  CHAR       xs         xs
BLOCK7  NUMERIC    arhs       arhs
BLOCK7  NUMERIC    birthyear  birthyear
/* record comparison same as for pass 3 above */
…
;

/*Pass 8: compare all records with the date of birth and same phonetic encoding of
suburb/locality/town*/
BLOCK8  NUMERIC    birthd     birthd
BLOCK8  CHAR       xl         xl
```

```
/* record comparison same as for pass 3 above */
…
;

/* summed match (comparison) weight cut-offs for deemed non-matches, possible
matches and deemed matches – these were set heuristically by examining ppairs of
records in the critical match weight regions after each pass */
CUTOFF1 20.0 20.0 50.0
CUTOFF2 30.0 20.0 50.0
CUTOFF3 44.5 36.0 50.0
CUTOFF4 55.0 46.0 55.0
CUTOFF5 55.0 43.0 55.0
CUTOFF6 55.0 43.0 55.0
CUTOFF7 55.0 43.0 55.0
CUTOFF8 55.0 30.0 55.0
/* directives not to use inverse frequency weighting of match weights for these
variables as there are too many distinct values for the software to handle */
VARTYPE Birthd NOFREQ
VARTYPE mrn    NOFREQ
VARTYPE mrn    NOUPDATE
                                                    ;
```

REFERENCES

References

1 Richters J, Grulich AE, de Visser RO, Smith AM, Rissel CE. Sex in Australia: Contraceptive practices among a representative sample of women. Aust N Z J Public Health 2003; 27:210-6

[2] Abma J, Chandra A, Mosher W, Peterson L, Piccinino L. Fertility, family planning, and women's health: New data from the 1995 National Survey of Family Growth. National Center for Health Statistics. Vital Health Stat 1997; 23(19)

[3] Ali MM, Cleland J, Shah IH. Condom use within marriage: a neglected HIV intervention. Bull World Health Organ. 2004 Mar;82(3):180-6.

[4] NHS Public Health Resource Unit. Critical Appraisal Skills Programme Appraisal Tools. Milton Keynes Primary Care Trust, 2002. Available at http://www.phru.nhs.uk/casp/appraisa.htm

[5] National Center for Biotechnology Information. PubMed Entrez (web site) Available at http://www.ncbi.nlm.nih.gov/

[6] Ebbert JO, Dupras DM, Erwin PJ. Searching the medical literature using PubMed: a tutorial. Mayo Clin Proc. 2003 Jan;78(1):87-91

[7] Kuper H, Nicholson A, Hemingway H. Searching for observational studies: what does citation tracking add to PubMed? A case study in depression and coronary heart disease. BMC Medical Research Methodology 2006, 6**:**4 [http://www.biomedcentral.com/1471-2288/6/4/abstract]

[8] Moawad AH, Hafez ESE. The oviduct and egg transport. In: Hafez ESE, editor. Human reproduction, conception and contraception. 2nd ed. Hagerstown (MD): Harper & Row; 1980. pp. 201-220

[9] Moore JG. Female Reproductive Anatomy. In: Hacker NF, Moore JG, editors. Essentials of Obstetrics and Gynecology. 3rd ed. Philadelphia: W.B. Saunders Company; 1998. pp. 3-11

[10] Lungren SS. A case of caesarean section twice successfully performed on the same patient: with remarks on the time, indications, and details of the operation. Am J Obstet Gynecol 1881; 14:76.

[11] Wheeless CR. Tubal sterilization. In: Thompson JD, Rock JA, editors. TeLinde's Operative Gynecology. Philadelphia: JB Lippincott; 1992. pp. 343 – –359

[12] Moore JG, DeCherney AH. Contraception and Sterilization. In: Hacker NF, Moore JG, editors. Essentials of Obstetrics and Gynecology. 3rd ed. Philadelphia: W.B. Saunders Company; 1998. p.p 516 – –531

[13] Evans TN. Female sterilization. In: Hafez ESE, editor. Human reproduction, conception and contraception. 2nd ed. Hagerstown (MD): Harper & Row; 1980. pp. 777–795

[14] Tatum HJ, Schimdt FH. Contraceptive and sterilization practices and extrauterine pregnancy: A realistic perspective. Fertil Steril 1977; 28:407-21

[15] Darabi KF, Richart RM. Collaborative study on hysteroscopic sterilization procedures: preliminary report. Obstet Gynecol 1977; 49:48-52.

[16] Cooper JM. Hysteroscopic sterilization. Clin Obstet Gynecol. 1992; 35:282-98.

[17] Neuwirth RS. Update on transcervical sterilization. Int J Gynaecol Obstet 1995; 51:S523-8

[18] Trussel TJ, Faden R, Hatcher RA. Efficacy information in contraceptive counselling: Those little white lies. AJPH 1976. 66(8):761-767

[19] Farley TMM. Life-table methods for contraceptive research. Stat Med 1986; 5:475-89

[20] Trussel J. Methodological pitfalls in the analysis of contraceptive failure. Stat Med 1991; 10:201-20

[21] Dieckmann WJ, Hauser EB. Pregnancy following sterilization. Am J Obstet Gynec 1948; 55:488.

[22] Lipscomb GH, Spellman JR, Ling FW. The effect of same-day pregnancy testing on the incidence of luteal phase pregnancy. Obstet Gynecol 1993; 82: 411-3

[23] Hosmer DW, Lemeshow, S. Applied Survival Analysis: Regression Modeling of Time to Event Data. New York: John Wiley & Sons; 1999. p18

[24] Pearl R. Factors in human fertility and their statistical evaluation. Lancet 1933; 2:607-11

[25] Berkson J, Gage RP. Calculation of survival rates for cancer. Proc Staff Meet Mayo Clin 1950; 25:270-286

[26] Potter RG. The multiple decrement life table as an approach to the measurement of use effectiveness and demographic effectiveness of contraception. Contributed papers, Sydney Conference International Union for the Scientific Study of Population, 1967. pp869-883

[27] Armitage P, Berry G. Statistical Methods in medical research. 2nd Ed. Oxford: Blackwell Scientific Publications, 1987. p426

[28] Peterson HB, Xia Z, Hughes JM, Wilcox LS, Tylor LR, Trussel J. The risk of pregnancy after tubal sterilization: Findings from the US Collaborative Review of Sterilization. Am J Obset Gynecol 1996 4(4):1161-1168.

[29] Greenwood M. The natural duration of cancer. Reports of Public Health and Medical Subjects, Vol 33, HMSO, London, 1926.

[30] Kalbfleish JD, Prentice RL. The Statistical Analysis of Failure Time Data. New York: Wiley; 1980.

[31] Harrell FE. Regression modelling strategies: with applications to linear models, logistic regression and survival analysis. New York: Springer-Verlag, New York, 2001. p401

[32] Rothman KJ, Greenland, S. Modern Epidemiology. 2nd Ed. Philadelphia: Lippincott-Raven, 1998. p37

[33] Garb AE. A review of tubal sterilization failures. Obstet Gynecol Surv 1957; 12:291-6

[34] Prystowsky H, Eastman, NJ. Puerperal tubal sterilization. Report of 1,830 cases. JAMA 1955, 158:463-7

[35] McElin TW, Buckingham JC, Johnson RE. Tubal sterilization. Study at Evanston Hospital. Am J Obstet Gynecol. 1967; 97:479-87

[36] Poulson AM Jr. Analysis of female sterilization technics. Obstet Gynecol. 1973; 42:131-5.

[37] Shah A, Courey NG, Cunnanan RG. Pregnancy following laparoscopic tubal electrocoagulation and division. Am J Obstet Gynecol 1977; 129:459-60

[38] Cheng MCE, Wong YM, Rochat RW, Ratnam SS. Sterilization failures in Singapore: an examination of ligation techniques and failure rates. Stud Fam Plann 1977 8:109-12

[39] Hughes GJ. Sterilisation failure. Br Med J 1977; 2:1337-9

[40] Keeping JD, Chang A, Morrison J. Sterilization: A Comparative review. Aust NZ J Obstet Gynaec 1979; 19:193-202

[41] Mumford SD, Bhiwandiwala PP, Chi IC. Laparoscopic and minilaparotomy female sterilisation compared in 15,167 cases. Lancet 1980; 2(8203):1066-70

[42] Mumford SD, Bhiwandiwala PP. Tubal ring sterilization: experience with 10,086 cases. Obstet Gynecol 1981; 57(2):150-7

[43] Bhiwandiwala PP, Mumford SD, Feldblum PJ. A comparison of different laparoscopic sterilization occlusion techniques in 24,439 procedures. Am J Obstet Gynecol 1982; 144(3):319-31

[44] Chi IC, Mumford SD, Gardner SD. Pregnancy risk following laparoscopic sterilization in non-gravid and gravid women. J Reprod Med 1981; 26:289-94

[45] Chi IC, Siemens AJ, Champion CB, Gates D, Cilenti D. Pregnancy following minilaparotomy tubal sterilization: An update of an international data set. Contraception 1987; 35:171-178.

[46] Chi IC, Laufe LE, Gardner SD, Tolbert MA. An epidemiologic study of risk factors associated with pregnancy following female sterilization. Am J Obstet Gynecol 1980; 136(6):768-773

[47] Rothman KJ, Greenland, S. Modern Epidemiology. 2nd Ed. Philadelphia: Lippincott-Raven, 1998. p110

[48] Lee NC, Rubin GL. Report of a high pregnancy rate after sterilization with the Bleier clip. South Med J 1984; 77:601-2

[49] Ayers JWT, Johnson RS, Ansbacher R, Menon M, LaFeria JJ, Roberts JA. Sterilization failures with bipolar tubal cautery. Fertile Steril 1984; 42:526-30

[50] Gunston KD, VanCoeverden Devan Coeverden de Groot HA, Bromhall MR. Pregnancy after tubal occlusion: a 5 year study. S afrAfr Med J 1983; 63:517-21.

[51] Vessey M, Huggins G, Lawless M, McPherson K, Yeates D. Tubal sterilization: Findings in a large prospective study. Br J Obstet Gynaecol 1983; 90:203-9

[52] Sitompul H, Lun KC, Lumbanraja M, Kaban RM, Albar E, Simanjuntak P, Hanafiah MJ. Comparison of three types of tubal sterilisation: The Medan experience. Contraception. 1984 Jan;29(1):55-63.

[53] Sherman PA, Burigo JA. Comparison of laparoscopic Falope-ring and minilaparotomy sterilization. Obstet Gynecol 1984; 63:71-75

[54] Aranda C, de Badia D, Mahran M, Feldblum PJ. A comparative clinical trial of the tubal ring versus the Rocket clip for female sterilization. Am J Obstet Gynecol. 1985; 153:755-9.

[55] Indian Council for Medical Research. Tubal sterilization with Filshie clip. Contraception 1984; 30:339-353

[56] De Villiers VP. Postpartum sterilisation with the Filshie titanium silicone-rubber clip and subsequent pregnancy. S Afr Med J 1987; 71:498-9

[57] Chi IC, Gates D, Bunce S, Rivera R, Apelo R, Ramos R, de la Vega JL. Timing of postpartum tubal sterilization using the Filshie clips: an analysis of data from two developing-country centers. Contraception. 1991; 43:33-44.

[58] De Villiers VP. Failed postpartum sterilisation - a comparison of 5 methods. S Afr Med J. 1992 Nov;82(5):355-6

[59] Yan JS, Hsu J, Yin CS. Comparative study of Filshie clip and Pomeroy method for postpartum sterilization. Int J Gynecol Obstet 1990; 33:263-267.

[60] Stovall TG, Ling FW, Henry GM, Ryan GM. Method failures of laparoscopic tubal sterilization in a residency training program: a comparison of the tubal ring and spring-loaded clip. J Reprod Med 1991; 36:283-6

[61] Birdsall MA, Pattison NS, Wilson P. Female sterilisation: National Women's Hospital 1988-9. N Z Med J. 1994;107:473-5.

[62] Makar AP, Vanderheyden JS, Schatterman EA, Albertyn GP, Verkinderen JJ, Van Marck EA. Female sterilization failure after bipolar electrocoagulation: a 6 year retrospective study. Eur J Obstet Gynecol Reprod Biol 1990; 37:237-46

[63] Trias FM, Anderson JE, Ojeda G, Oberle MW. A lifetable analysis of sterilization failure: Data from the Profamilia Clinic, Bogota, Columbia. Int J Gynaecol Obstet 1987; 25:235-40

[64] Lassner KJ, Chen CHC, Oberle MW, da Trindade TCSM Aguinaga H. Analysis of sterilization failure in Brazil. Int J Gynecol Onstet 1988; 27:255-263.

[65] Peterson HB, Xia Z, Hughes JM, Wilcox LS, Tylor LR, Trussell J. The risk of pregnancy after tubal sterilization: Findings from the U.S. Collaborative Review of Sterilization. Am J Obstet Gynecol 1996; 174:1161-1168.

[66] Loffer FD, Pent D. Pregnancy after laparoscopic sterilization. Obstet Gynecol 1980; 55:643-648

[67] Hulka FJ. Spring clip sterilization: One-year follow-up of 1,000 cases. Proceedings of a Workshop on Advances in femaleFemale Sterilization Techniques. Sciarra JJ, Droegmueller W, Speidel JJ (eds.). Hagerstown, MD: Harper and Row, 1975. p51

[68] Loffer FD, Pent D. Indications, contraindications and complications of laparoscopy. Obstet Gynecol Surv 1975; 30:407-13

[69] Chi IC, Feinblum I. Luteal phase pregnancies in female sterilization patients. Contraception 1981; 21:579-89

[70] Brenner PF, Benedetti T, Mishell DR. Ectopic pregnancy following tubal sterilization surgery. Obstet Gynecol 1977; 49:323-4

[71] Chakravarti S, Shardlow J. Tubal pregnancy after sterilization. Br J Obstet Gynaecol 1975; 82:58-60

[72] Sivanesaratnam V, Ng KH. Tubal pregnancies following postpartum sterilization. Fertil Steril 1976l 26:945

[73] Napolitano PG, Vu K, Rosa C. Pregnancy after failed tubal sterilization. J Reprod Med 1996; 41:609-613

[74] Greisman B. Ectopic pregnancy in women with previous tubal sterilizations at a Canadian community hospital. J Reprod Med 1991; 36:206-209

[75] Stock RJ, Nelson KJ. Ectopic pregnancy subsequent to sterilization: Histologic evaluation and clinical implications. Fertil Steril 1984; 42:211-15

[76] Wolf GC, Thompson NJ. Female sterilization and subsequent ectopic pregnancy. Obstet Gynecol 1980; 55:17-19

[77] McCausland A. High rate of ectopic pregnancy following laparoscopic tubal coagulation failures: incidence and etiology. Am J Obstet Gynecol 1980; 136:97-101

[78] Hughes GJ. Ectopic pregnancy after sterilisation. Med J Aust. 1980; 1:275.

[79] Metz KGP, Mastroianni L. Tubal pregnancy subsequent to transperitoneal migration of spermatozoa. Obstet Gynecol Surv 1979; 34:554-60

[80] Brenner PF, Benedetti T, Mishell DR. Ectopic pregnancy following tubal sterilization surgery. Obstet Gynecol 1977; 49:323-4

[81] Lubell I, Neuwirth RS, Frisher R. A hidden risk: Extrauterine pregnancy after tubal sterilization. Adv Plann Parent 1978; 13:24-9

[82] Hornstein S, Kay SA. Abdominal pregnancy following Irving tubal ligation. Obstet Gynecol 1959; 13:337-340

[83] Kjer JJ, Knudsen LB. Ectopic pregnancy subsequent to laparoscopic sterilization. Am J Obstet Gynecol 1989; 160:1202-4

[84] Shah JP, Parulekar SV, Hinduia IN. Ectopic pregnancy after tubal sterilization. J Postgrad Med 1990; 37:17-20

[85] Davis MR. Recurrent ectopic pregnancy after tubal sterilization. Obstet Gynecol 1986; 68:445-6

[86] Breen JL. A 21 year survey of 654 ectopic pregnancies. Am J Obstet Gynecol 1970; 106:1004-9.

[87] Honore LH, O'Hara KE. Failed tubal sterilization as an etiological factor in ectopic tubal pregnancy. Fertil Steril 1978; 29:509-11

[88] DeStefano F, Peterson HB, Layde PM, Rubin GL. Risk of ectopic pregnancy following tubal sterilization. Obstet Gynecol 1982; 60:326-330

[89] Vessey M, Doll R, Peto R, Johnson B, Wiggins P. A long-term follow-up study of women using different methods of contraception – An interim report. J Biosoc Sci 1976; 8:373.

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

[91] Westrom L, Bengtsson LPH, Mardh PA. Incidence, trends and risks of ectopic pregnancy in a population of women. Br Med J (Clin Res Ed). 1981;282:15-18.

[92] 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 Reprod Fertil. 1985; 3:131-43.

[93] Marchbanks PA, Annegers JF, Coulam CB, Strathy JH, Kurland LT. Risk factors for ectopic pregnancy: a population based study. JAMA 1988; 259:1823-1827

[94] Holt VL, Chu J, Daling JR, Stergachis AS, Weiss NS. Tubal sterilization and subsequent ectopic pregnancy: A case-control study. JAMA 1991: 266:242-246

[95] Peterson HB, Xia Z, Hughes JM, Wilcox LS, Taylor LR, Trussell J. The risk of ectopic pregnancy after tubal sterilization. N Engl J Med 1997; 336:762-7

[96] Hendrix NW, Chauhan SP, Maier RC. Ectopic pregnancy in sterilized and nonsterilized women: A comparison. J Reprod Med 1998; 43:515-520

[97] Templeton AA, Cole S. Hysterectomy following sterilisation. Br J Obstet Gynaecol 1982; 89:845-8

[98] Heasman MA, Clark JA. Medical record linkage in Scotland. Health Bulletin 1979; 37:97-103.

[99] Cohen MM. Long-term risk of hysterectomy after tubal sterilization. Am J Epidemiol 1987; 125:410-9

[100] Golhaber MK, Armstrong MA, Golditch IM, Sheehe PR, Petitti DB, Friedman GD. Long-term risk of hysterectomy among 80,007 sterilized and comparison women at Kaiser Permanente, 1971-1987. Am J Epidemiol 1993; 138:508-21

[101] Stergachis A, Shy KK, Grothaus C, Wagner EH, Hecht JA, Anderson G. Tubal sterilization and the long-term risk of hysterectomy. JAMA 1990;264:2893-8

[102] Hillis SD, Marchbanks PA, Taylor LR, Peterson HB. Tubal sterilization and long-term risks of hysterectomy: Finding from the US Collaborative Review of Sterilization. Obstet Gynecol 1997; 89:609-14

[103] DeStefano F, Huezo CM, Peterson HB, Rubin GL, Layde PM, Ory HW. Menstrual changes after tubal sterilization. Obstet Gynecol 1983; 62:673-81

[104] Rulin MC, Davidson AR, Philliber SG, Graves WL, Cushman LF. Changes in menstrual symptoms among sterilized and comparison women: A prospective study. Obstet Gynecol 1989; 74:145-54

[105] Foulkes J, Chamberlain G. Effects of sterilization on menstruation. South Med J 1985; 78:544-7

[106] Bhiwandiwala PP, Mumford SD, Feldblum PF. Menstrual pattern changes following laparoscopic sterilization with different occlusion techniques: A review of 10,004 cases. Am J Obstet Gynecol 1983; 145:684-94

[107] Wilcox LS, Martinez-Schnell B, Peterson HB, Ware JH, Hughes JM. Menstrual function after tubal sterilization. Am J Epidemiol 1992; 135:1368-81

[108] DeStefano F, Perlman JA, Peterson HB, Diamond EL. Long-term risk of menstrual disturbances after tubal sterilization. Am J Obstet Gynecol 1985; 152:835-41

[109] Lawson S, Cole RA, Templeton AA. The effect of laparoscopic sterilisation by diathermy or silastic bands on post-operative pain, menstrual symptoms and sexuality. Br J Obstet Gynaecol. 1979;86:659-663

[110] Alvarez F, Tejada AS, Faundes A, Segal S, Brache V. Prospective study of the pituitary-ovarian function after tubal sterilization by the Pomeroy or Uchida techniques. Fertil Steril 1989; 51:604-8.

[111] Donnez J, Wauters M, Thomas K. Luteal function after tubal sterilization. Obstet Gynecol 1981; 57:65-8.

[112] Peterson HB, DeStefano F, Rubin GL, Greenspan Jr, Lee NC, Ory HW. Deaths attributable to tubal sterilization in the United States, 1977 to 1981. Am J Obstet Gynecol 1983; 146:131-6

[113] Rochat RW, Bhiwandiwala PP, Feldbum PJ, Peterson HB. Mortality associated with sterilization: preliminary results of an international collaborative observational study. Int J Gynaecol Obstet 1986; 24:275-284.

[114] Peterson HB, DeStefano F, Greenspan JR, Ory HW. Mortality risk associated with tubal sterilization in United States hospitals. Am J Obstet Gynecol. 1982 May 15;143(2):125-9.

[115] Destefano F, Greenspan JR, Dicker RC, Peterson HB, Strauss LT, Rubin GL. Complications of interval laparoscopic tubal sterilization. Obstet Gynecol. 1983; 61:153-8.

[116] Peterson HB, Greenspan JR, Ory HW, DeStefano F. Tubal sterilization mortality surveillance, United States, 1978-1979. Adv Plann Parenth 1981; 16:71-73.

[117] Australian Institute of Health and Welfare (AIHW). GRIM (General Record of Incidence of Mortality) Books. AIHW: Canberra, 2005.

[118] Rubin GL, Ory HW, Layde PM. The mortality risk of voluntary surgical contraception. Biomed Bull. 1982; 3:1-5.

[119] Westhoff C, Davis A. Tubal sterilization: focus on the U.S. experience. Fertil Steril. 2000; 73:913-22.

[120] Rock JA, Guzick DS, Katz E, Zacur HA, King TM. Tubal anastomosis: pregnancy success following reversal of Falope-ring or monopolar sterilization. Fertile Steril 1987; 48:13-17.

[121] Henderson SR. The reversibility of female sterilization with the use of microsurgery: a report on 102 patients with more than one year of follow-up. Am J Obstet Gynecol 1984; 149:57-65.

[122] Spivak MM, Liprach CL, Rosenthal DM. Microsurgical reversal of sterilization: a six-year study. Am J Obstet Gynecol 1986; 154:355-61.

[123] DeCherney AH, Mezer HC, Naftolin F. Analysis of failure of microsurgical anastomosis after midsegment non-coagulation tubal ligation. Fertile Steril 1983; 39:618-22.

[124] Hulka J, Noble A, Letchworth A, Lieberman B, Owen E, Gomel V. Reversibility of clip sterilizations (letter). Lancet 1982; 2(8304):927.

[125] Grubb GS, Peterson HB, Layde PM, Rubin GL. Regret after decision to have a tubal sterilization. Fertil Steril. 1985; 44:248-53.

[126] Hillis SA, Marckbanks PA, Taylor LR, Peterson HB. Post-sterilization regret: Findings from the United States Collaborative Review of Sterilization. Obstet Gynecol 1999; 93:889-95.

[127] Nardin JM, Kulier R, Boulvain M. Techniques for the interruption of tubal patency for female sterilisation. Cochrane Database Syst Rev. 2003; CD003034.

[128] Kulier R, Boulvain M, Walker D, De Candolle G, Campana A. Minilaparotomy and endoscopic techniques for tubal sterilisation. Cochrane Database Syst Rev 2004; CD001328

[129] Argent VP. Failed female sterilization and the law. Med Sci Law. 1985; 25:136-42.

[130] Alvey W, Jamerson B (eds). Record Linkage techniques – 1997. Proceedings of an International Workshop and Exposition. Washington, DC: Federal Committee on Statistical Methodology, Office of Management and Budget, 1997.

[131] Anderson RJ. Information technology in medical practice: safety and privacy lessons from the United Kingdom. Med J Aust 1999; 170:181-184

[132] Newcombe HB, Kennedy JM, Axford SJ, James AP. Automatic Linkage of Vital Records. Science 1959; 130:954-959

[133] Newcombe HB, Kennedy JM. Record Linkage: Making Maximum Use of the Discriminating Power of Identifying Information. Communications of the Association for Computing Machinery 1962; 5:563-566.

[134] Wajda A, Roos LL, Layefsky M, Singleton JA. Record Linkage Strategies: Part II. Portable Software and Deterministic Matching. Meth Inform Med 1991; 30:210-4.

[135] Fellegi IP, Sunter AB. A Theory for Record Linkage. Journal of the American Statistical Society 1969; 40:1183-1210.

[136] Winkler WE. Using the EM algorithm for weight computation in the Fellegi-Sunter model of record linkage. Research Report RR 2000-05. Washington, DC: US Bureau of the Census, 2000

[137] Winkler WE. Frequency-based matching in the Fellegi-Sunter Model of Record Linkage. Research Report RR 2000-06. Washington, DC: US Bureau of the Census, 2000

[138] Jaro MA. Probabilistic linkage of large public health data files. Stat Med. 1995; 14:491-8.

[139] Kelley RP. Blocking considerations for record linkage under conditions of uncertainty. In: Proceedings of the Social Statistics Section, American Statistical Association, 1984. pp602-5

[140] AutoStan and AutoMatch, User's Manuals. Kennebunk, Maine: MatchWare Technologies Inc., 1998.

[141] McCallum AK, Nigam K, Ungar L. Efficient clustering of high-dimensional data sets with application to reference matching. In: Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining (KDD-2000), Boston, MA, 2000. New York: Association for Computing Machinery; 2001.

[142] Baxter R, Christen P, Churches T. A Comparison of Fast Blocking Methods for Record Linkage. In: Proceedings of the ACM Workshop on Data Cleaning, Record Linkage and Object Identification, Washington, DC, August 2003. New York: Association for Computing Machinery; 2004.

[143] Churches T, Christen P, Lim K, Zhu JX. Preparation of name and address data for record linkage using hidden Markov models. BMC Med Inform Decis Mak. 2002; 2:9. Available at http://www.biomedcentral.com/1472-6947/2/9

[144] Newcombe HB, Fair ME, Lalonde P. The use of names for linking personal records. J Amer Stat Assoc 1992;87:1193-204.

[145] Arrellano MG. Comment on: Newcombe HB, Fair ME, Lalonde P. The use of names for linking personal records. J Amer Stat Assoc 1992; 87:1204-6.

[146] Christen P, Churches T. Febrl - Freely extensible biomedical record linkage. ANU Computer Science Technical Reports TR-CS-02-05. Canberra: Australian National University, 2002.

[147] Roos LL, Wadja A. Record linkage strategies. Part 1: Estimating information and evaluating approaches. Meth Inform Med 1991; 30:117-23

[148] Holman CDJ. Introductory Analysis of Linked Health Data: Principles and Hands-On Applications, Short Course Workbook, Version 1.2, January 2007. Sydney: Health Evaluation Research and Outcomes Network (HERON) and School of Population Health, University of Western Australia. Ch. 2 p6

[149] Hernandez MA, Stolfo SJ. The Merge/Purge Problem for Large Databases, in Proceedings of the SIGMOD Conference, San Jose, 1995.

[150] McGeechan K, Kricker A, Armstrong B, Stubbs J. Evaluation of linked cancer registry and hospital records of breast cancer. Aust N Z J Public Health. 1998; 22:765-70.

[151] Adelson P, Lim K, Churches T, Nguyen R. Surgical treatment of breast cancer in New South Wales 1991, 1992. Aust N Z J Surg. 1997; 67:9-14.

[152] Churches T, Lim K. Using record linkage to measure trends in breast cancer surgery. N S W Public Health Bull. 2001; 12:105-110.

[153] Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. Int J Epidem 2002; 31:1246-52.

[154] Belin TR, Rubin DB. A method for calibrating false-match rates in record linkage. Journal of the American Statistical Association 1995; 90:694-707.

[155] Scheuren F, Winkler WE. Regression analysis of data files that are computer matched – part 1. Survey Methodology 1993; 19:39-58.

[156] Scheuren F, Winkler WE. Regression analysis of data files that are computer matched – part 2. Survey Methodology 1997; 23:157-165.

[157] Winkler WE. Methods for Record Linkage and Bayesian networks. Research Report RR 2002-05. Washington, DC: US Bureau of the Census, 2002.

[158] Gu L, Baxter RA, Vickers D, Rainsford C. Record Linkage: Current Practice and Future Directions. CSIRO Mathematical and Information Sciences Technical Report 03/83. Canberra: CSIRO, 2003.

[159] Borthwick A, Buechi M, Goldberg A. Key Concepts in the ChoiceMaker 2 Record Matching System. In: Proceedings of the KDD-2003 Workshop on Data Cleaning, Record Linkage, and Object Consolidation, 2003, Washington, DC, pp28-30

[160] Verykios VS, Elmagarmid AK, Houstis EN. Automating the Approximate Record Matching Process. Journal of Information Sciences 2000; 126:83-98.

[161] Boruch R, Cecil J. Assuring the Confidentiality of Social Research Data. Philadelphia: University of Philadelphia Press, 1979.

[162] Pommerening K, Miller M, Schidtmann I, Michaelis J. Pseudonyms for cancer registries. Methods Inf Med 1996, 35:112-121.

[163] Kelman CW, Bass AJ, Holman CD. Research use of linked health data- - a best practice protocol. Aust N Z J Public Health 2002, 26**:**251-255.

[164] Churches T. A proposed architecture and method of operation for improving the protection of privacy and confidentiality in disease registers. BMC Medical Research Methodology 2003, 3:1 (available at http://www.biomedcentral.com/1471-2288/3/1)

[165] Dusserre L, Quantin C, Bouzelat H. A one way public key cryptosystem for the linkage of nominal files in epidemiological studies. Medinfo 1995, 8:644-7.

[166] Quantin C, Bouzelat H, Dusserre L. A computerized record hash coding and linkage procedure to warrant epidemiological follow-up data security. Stud Health Technol Inform 1997, 43:339-42.

[167] Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, Dusserre L. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. Methods Inf Med 1998, 37**:**271-7.

[168] Agrawal R, Evfimievski A, Srikant R. Information sharing across private databases. In: Proceedings of the 2003 ACM SIGMOD international conference on Management of data, San Diego, 2003. New York, 2004: Association for Computing Machinery. pp 86-97.

[169] Churches T, Christen P. Some methods for blindfolded record linkage. BMC Medical Informatics and Decision Making 2004, 4:9 (available at http://www.biomedcentral.com/1472-6947/4/9)

[170] O'Keefe C, Yung M, Gu L, Baxter R. Privacy-preserving data linkage protocols. In: Proceedings of the 2004 ACM workshop on Privacy in the electronic society,

Washington, DC, 2004. New York, 2004: Association for Computing Machinery. pp 94-102

[171] NSW Department of Health. Information Privacy Code of Practice, Second Edition. North Sydney, May 1996: NSW Department of Health.

[172] New South Wales Government Gazette. Health Records Information Privacy Act 1992. Sydney, 2002: NSW Government Printing Service. (Full text available at http://www.austlii.edu.au/au/legis/nsw/consol_act/hraipa2002370/_).

[173] Privacy NSW. Health Records and Information Privacy Act 1992 (NSW): Statutory guidelines on research. Sydney, 2004: privacy NSW. (Available at http://www.lawlink.nsw.gov.au/lawlink/privacynsw/ll_pnsw.nsf/pages/PNSW_03_hripact#4b)

[174] Free Software Foundation. The GNU Privacy Guard (GPG) [software]. Available http://www.gnupg.org/

[175] New South Wales Department of Health. NSW Health Data Collections: Inpatient Statistics Collection (web site). Available at http://www.health.nsw.gov.au/im/ims/isc/

[176] Muscatello D, Travis S. Using the International Classification of Diseases with HOIST. N S W Public Health Bull. 2001 12:289-293

[177] Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. Aust N Z J Public Health. 1999 23:453-9.

[178] The Sax Institute. The 45 and Up Study. [web site] Available at http://www.45andup.org.au/

[179] Shankar M, Economides DL, Sabin CA, Tan B, Kadir RA. Outpatient medical management of missed miscarriage using misoprostol. J Obstet Gynaecol 2007; 27:283-6.

[180] SAS Institute. The SAS System Version 8.2 for Microsoft Windows. Cary, NC, USA, 2001: The SAS Institute.

[181] National Centre for Classification in Health. ICD-9-CM manual, 1st, 2nd and 3rd Eds. Sydney: National Centre for Classification in Health, 1994-98.

[182] National Centre for Classification in Health. ICD-10-AM manual, 1st Ed. Sydney: National Centre for Classification in Health, 1999.

[183] Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care 1994; 32:81-90.

[184] Henderson T, Shepheard J, Sundararajan V. Quality of diagnosis and procedure coding in ICD-10 administrative data. Med Care 2006; 44:1011-9.

[185] MacIntyre CR, Ackland MJ, Chandraraj EJ, Pilla JE. Accuracy of ICD-9_CM codes in hospital morbidity data, Victoria: Implications for public health research. Aust N Z J Public Health 1997; 21:477-82.

[186] Jorm LR. The Centre for Health Record Linkage: Linking health data for NSW and the ACT [workshop presentation] Sydney, 2006: Sax Institute. Available at http://www.saxinstitute.org.au/contentUploadedByEWeb/Files/Louisa%20Jorm%20workshop%206%20Nov%202006%2Epdf

[187] Bell M. How often do Australian move? Alternative measures of population mobility. Journal of the Australian Population Association, 1996; 13:101-124

[188] R Development Core Team (2006). R: A language and environment for statistical computing. Vienna, Austria, 2006: R Foundation for Statistical Computing, ISBN 3-900051-07-0. Available at http://www.R-project.org.

[189] Public Health Division. The health of the people of New South Wales - Report of the Chief Health Officer. Sydney, 2006: NSW Department of Health. Available at: http://www.health.nsw.gov.au/public-health/chorep/toc/index.htm

[190] Tukey JW. Exploratory Data Analysis, Boston, MA, 1977: Addison-Wesley. pp56-73

[191] S original by Terry Therneau and ported by Thomas Lumley. survival: Survival analysis, including penalised likelihood.. R package version 2.31. Available at http://www.r-project.org/cran/

[192] Mueller HG, Wang JL. Hazard rates estimation under random censoring with varying kernels and bandwidths. Biometrics 1994, 50:61-76

[193] Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. Biometrika 1982; 69**:**553-566.

[194] Fox J. An R and S-PLUS Companion to Applied Regression. London: Sage Publications, 2002.

[195] Royston P, Altman D. Regression using fractional polynomials of continuous covariates. Appl Stat. 1994; 3:429–467.