FERTILITY IN POPULATIONS AND IN PATIENTS

Population studies on natural fertility and prediction of treatment outcome in anovulatory infertile patients

René Eijkemans

Publication of this thesis was financially supported by FERRING BV, NV Organon, Serono Benelux BV and the Department of Public Health, ErasmusMC, University Medical Center Rotterdam, who are hereby gratefully acknowledged for their contribution.

FERTILITY IN POPULATIONS AND IN PATIENTS: Population studies on natural fertility and prediction of treatment outcome in anovulatory infertile patients / M.J.C. Eijkemans
Thesis Erasmus University Rotterdam

ISBN 90-9018399-X

Cover photograph: © Eric Wanders/www.fotonatura.com
Printed by PrintPartners Ipskamp, Enschede

FERTILITY IN POPULATIONS AND IN PATIENTS

Population studies on natural fertility and prediction of treatment outcome in anovulatory infertile patients

VRUCHTBAARHEID IN POPULATIES EN IN PATIËNTEN

Populatiestudies naar natuurlijke vruchtbaarheid en predictie van de resultaten van behandeling van patiënten met anovulatoire infertiliteit

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof.dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 15 september 2004 om 13.45 uur

door

MARINUS JOHANNES CORNELIS EIJKEMANS

geboren te St. Michiels Gestel.

PROMOTIECOMMISSIE

Promotoren:
Prof.dr. J.D.F. Habbema
Prof.dr. E.R. te Velde
Prof.dr. B.C.J.M. Fauser

Overige leden:
Prof.dr. Th. Stijnen
Prof.dr. Th.J.M. Helmerhorst
Prof.dr. J. Collins

Contents

General Discussion

Chapters 2, 3, 4, 6, 7, 8, 9 and 10 of this PhD thesis are based on manuscripts that have been published previously or have been submitted to a journal. Published manuscripts have been reproduced with explicit permission from the publishers.

2. Eijkemans MJ, Habbema HD, te Velde ER. The end of female fertility is highly variable and related to fertility at young age. *submitted.*
3. Eijkemans MJ, Habbema HD, Leridon H, te Velde ER. Primary and secondary infertility before the era of birth control: lessons from the past. *submitted.*
4. te Velde ER, Eijkemans R, Habbema HD. Variation in couple fecundity and time to pregnancy, an essential concept in human reproduction. *Lancet.* 2000;355:1928-1929.
6. Eijkemans MJ, Habbema JD, Fauser BC. Characteristics of the best prognostic evidence: an example on prediction of outcome after clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *Semin Reprod Med.* Feb 2003;21:39-47.
7. Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 1998;83:2361-2365.
8. Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. Predictors of chances to conceive in ovulatory patients during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 1999;84:1617-1622.
9. Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. A nomogram to predict the probability of live birth after clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *Fertil Steril.* 2002;77:91-97.
10. Eijkemans MJ, Imani B, Mulders AG, Habbema JD, Fauser BC. High singleton live birth rate following classical ovulation induction in normogonadotrophic anovulatory infertility (WHO 2). *Hum Reprod.* Nov 2003;18:2357-2362.

Preface

Procreation is a fact of life; the ability to have offspring may be considered to be *the* feature that defines 'life' itself, as it is a necessary condition for survival of the species. For individuals, this is not so self-evident. People may choose to remain childless, and the ones who wish to have children will not always succeed. For such an essential biological treat, the natural variation in fertility between individuals is huge: the monthly chance to become pregnant and have a child is around 0.2 on average, but it varies between couples from values as high as 0.5 to very low values, and for some couples the chance is effectively zero. While in previous times this just had to be accepted, modern Reproductive Medicine has developed treatment methods for couples that have an unresolved child wish. This doesn't mean that all of these couples will become pregnant: chance also plays a role in the success of treatment. This thesis is devoted to chances in fertility. It treats two subjects that are of interest to clinical Reproductive Medicine: in Part I, natural fertility chances in a population and in Part II, chances on outcome of treatment of one particular group of infertility patients.

Part I analyses a 19th century population. Two research questions are investigated with the aim to learn something from the past: one concerns the variation in the age at which female fertility comes to an end, and seeks evidence whether this variation might be related to the biological process of ovarian ageing. The second question concerns the pregnancy chances after a period of unsuccessful trying. This is of interest to clinicians who have to decide to start treatment in patients without an obvious diagnosis.

The research subject of Part II is the prediction of pregnancy chances after treatment of patients who have fertility problems due to cycle disturbances. The aim is to construct statistical models that may be used to predict -for individual patients- the pregnancy chances with the standard treatment protocol. These prediction models are then used in an analysis in which the efficiency of the treatment protocol is compared with several alternative protocols.

Part I

Population studies on natural fertility

1

Introduction part I: Lessons from a historical natural fertility population for modern reproductive medicine

BACKGROUND

In most western societies, reproductive behavior is nowadays controlled to a high degree: observed fertility patterns in a population -how many children do couples get and when do they get them- are the result of choices made by individuals rather than biological factors. This is in marked contrast with previous times, in which it was the biological destiny of women to have children – or to remain childless involuntarily. The driving force behind the change that has occurred is undoubtedly the availability of reliable contraception (Leridon, 1998), which has gone hand in hand with changes in social attitudes with respect to reproduction (e.g. secularization, emancipation of women (Blossfeld, 1995)). In many Western European societies, the number of children per woman has dropped below the natural replacement level (Von Cube, 1986), and everywhere couples who want to have children carefully plan when to have them. In many countries women delay childbearing to later age, the most pronounced example being the Netherlands in which the mean age at first childbirth has risen from 24.6 years in 1970 to 29.2 years in 2001. Since pregnancy chances decrease with age of the women (Schwartz and Mayaux, 1982) (van Noord-Zaadstra *et al.*, 1991), postponement of the first child has led to an increase in incidence of subfertility (te Velde, 1991).

Reproductive medicine has developed out of the need of couples that have an unresolved child wish. When this condition lasts for more than one year, couples are diagnosed as "infertile" and in most cases some form of treatment is started within 1 or 2 years following diagnosis unless the couple becomes pregnant spontaneously. Knowledge on natural fertility may help reproductive medicine both in the clinical question of deciding whether to treat or to wait and in research questions on the biology of human reproduction. A number of these questions cannot be answered in a

modern society because a) fertility is expressed in a highly controlled fashion and b) the natural course of "disease" is disturbed by treatment. Instead, data of natural populations could be used. Up till now, studies on fertility in natural populations have been mainly demographically or biometrically oriented (Wood, 1994) and studies with a clinical perspective have been rare (Potter and Parker, 1964) (Schwartz, 1981) (Menken *et al.*, 1986). In Part I of this thesis, a unique database on a large historical natural population from Quebec, Canada, is used to study two research questions from clinical reproductive medicine.

THE BALSAC DATABASE ON A HISTORICAL NATURAL FERTILITY POPULATION FROM QUEBEC, CANADA

Description of the population and the BALSAC database

The studies in this thesis use data on the population of the Saguenay Lac St-Jean (SLSJ) region in Quebec, Canada. The SLSJ region became populated by colonists from the nearby Charlevoix region from 1838 onwards. For a long time it was geographically isolated from the rest of Quebec (The highway to Quebec City was established only in 1951) (Bouchard, 1996). Due to their tradition of short lactation, the total number of children per woman was high in comparison with other natural populations (Lalou, 1990), which makes them ideally suited to study human fertility under natural conditions. The demographic fertility transition occurred relatively late in this region, in 1946-1955 among farmers and in 1931-1935 in others (Bouchard and Roy, 1991). The main reasons for this late transition lie in the fact that the population was predominantly rural and agricultural, little educated and culturally dominated by the Roman Catholic Church, leading to a paternalistic society (Bouchard, 2000). The really big change in fertility in the whole of Quebec occurred during the 'Quiet revolution' in the 1960's: Between 1959 and 1971, Quebec moved from the position of having the highest birth rate in Canada to that of the lowest (Bélanger, 2000).

Data of the population of the SLSJ region from church registers (Roman-Catholic) on baptism, marriage and death are contained in the 'BALSAC' demographic database at the Interuniversitary Institute for Population Research (IREP) in Chicoutimi, Quebec, Canada (Bouchard *et al.*, 1989). The database comprises the entire population of the region from the beginning of the colonisation in 1838 until 1970. We selected all children from N = 2,683 grandmothers who had their last child before 1900: 10,216 sons and 9,317 daughters, together with their (grand)children. The analyses described in Part I of this thesis are based on the data of the daughters.

Level of natural fertility

Demographic definitions

In demography, a clear distinction is made between the physiological capacity to get children, and the actual result of this capacity (Wood, 1994). The physiological

capacity of a woman, man, couple, group, or population to produce live children is called *fecundity*. The lack of that capacity is called *infecundity* or *sterility*. The probability to conceive during a month of unprotected intercourse is called *fecundability* (Gini, 1924). The actual reproductive performance is called *fertility*, and is used to denote the number of live born children during a certain period. *Infertility* refers to the absence of live births. Much demographic research has been focused on describing and explaining the variation in fertility both between populations and between individuals within a population. The main tool of understanding the variation has been to try to distinguish the various factors that determine fertility, and seek for factors that are constant across populations or individuals. Most importantly, the process of producing a live birth has been broken down into three components that depend on the age of a woman: the monthly probability of conceiving, the probability of fetal loss and the probability of being definitively sterile. However, many difficulties arise in such an approach and rather advanced computational problems have to be solved (Trussell and Wilson, 1985) (Barrett, 1986). In the literature on clinical reproductive medicine, a growing awareness is emerging that the final goal to achieve is a live birth (Collins *et al.*, 1995) (Snick *et al.*, 1997) (Vail and Gardener, 2003) (Barlow, 2003). Therefore, data on live births are well suited for our research questions that are inspired by clinical considerations.

When do we speak of natural fertility?

The classical definition of natural fertility was given by the French demographer Louis Henry (Henry, 1961): "By natural fertility we mean fertility which exists or has existed in the absence of deliberate birth control....Control may be said to exist when the behavior of the couple is bound to the number of children already born and is modified when this number reaches the maximum which the couple does not want to exceed.". This definition emphasizes two separate aspects of birth control: it should be deliberate and parity-dependent. Social factors such as taboos on pre-marital intercourse, sexual taboos during lactation or a prescribed period of lactation, limit the number of children, but without the couple intending to do so. Nor do they depend on the number of children already born. Therefore, natural fertility is not determined solely by physiological factors. Since our main interest is in physiological fertility we will have to devote some attention to behavioral factors, most notable the period of post-partum amenorrhea caused by breastfeeding.

Female age at marriage

Female age at first marriage was very young in this population: the mean was 21.3 years (SD = 5.1), the median was 20 years. In older pre-industrialization European populations, the mean ages at marriage ranged from 24.6 to 26.7 years (Dixon, 1971). The distribution is skewed to the right, as seen in **Figure 1**.

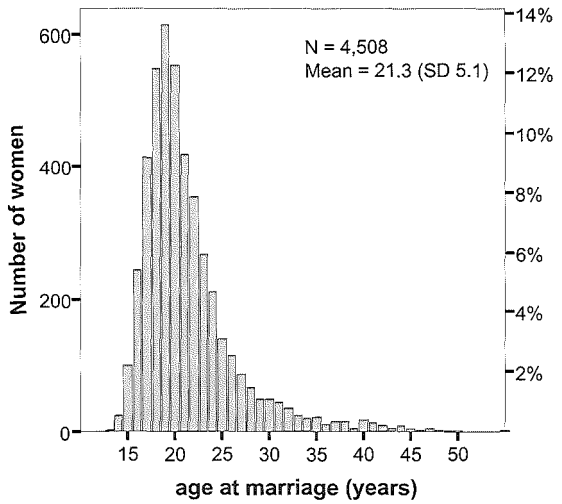


Figure 1 Distribution of female age at first marriage of 4,508 women born in the SLSJ region between 1840 and 1900.

The level of fertility

The level of fertility in this young marrying population is well expressed by the total number of children that women have. This number may be calculated for all married women, and for the subgroup of women who remain married at least until the age of 50. In a demographic sense, these are the women who have been able to complete their reproductive career. The distribution of number of children per woman is depicted in **Figure 2.** The mean number of children per woman was 7.9 (SD 4.8) and 9.1 (SD 4.7) respectively. The maximum number of children born to a woman was 22. This woman was born in 1890, married at age 16, had her last child at age 43 and died at age 68.

Demographic indicators for a limitation of fertility

We speak of natural fertility when no attempt is made to limit the family size by using birth control measures. The religious nature of the population (Roman Catholic) and the ongoing tradition of high fertility in the whole of Quebec must have put social pressure on couples to get children: the total fertility rate for the whole population of Quebec was 5.6 children per woman in 1891 (Henripin, 1968). According to Henry (Henry, 1976), the absence of birth control could be verified by comparing the age-specific number of children per married woman per year (or age-specific marital fertility rate) for different groups of age-at marriage. If couples used birth control to limit their family size, a difference in number of children per year would be noticed between women who married young and women who married at a later age. We have

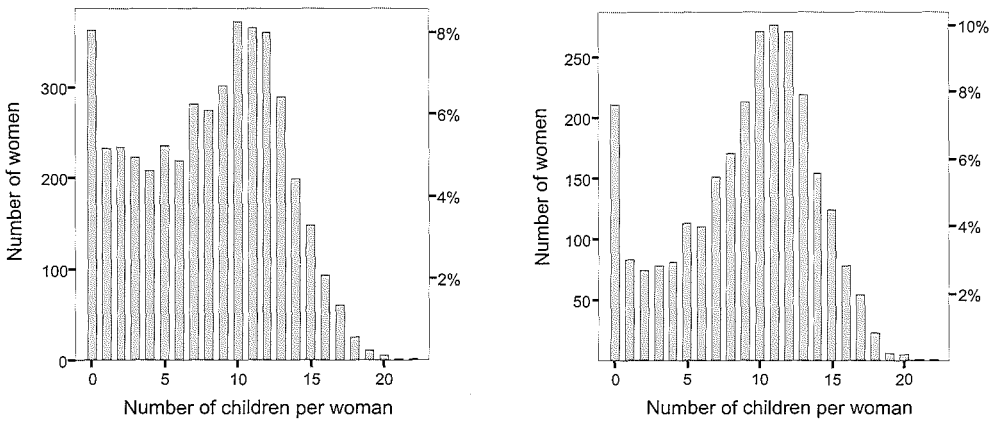


Figure 2 Distribution of number of children per women, for all women (left panel) and for the selection of women who were married until at least age 50 (right panel). Data from the SLSJ population.

performed such an analysis; the results are depicted in **Figure 3**. It is clear from this figure that there is no indication of limitation of family size in this population.

The high level of fertility had a severe negative impact on the life expectancy of women, as in pre-modern times maternal mortality was a major killer. This is clearly illustrated in **Table 1** where we look at the expected total duration of life at various ages. While the difference in expected total duration of life between birth and age 20 is

Figure 3 Marital fertility rate in 5-years classes of age of the woman for 6 groups of age at marriage. Data from the SLSJ population

Table 1. Expected total duration of life (years) at birth, and when reaching ages 20 and 50. Men and women from the SLSJ population, born in 1840-1899, and from the Netherlands in 1998 for comparison.

	SLSJ, Quebec		Netherlands 1998	
	Men	Women	Men	Women
At birth	42.9	40.8	74.8	80.5
At age 20	65.0	60.3	75.2	80.6
At age 50	73.2	73.0	76.7	81.7

a reflection of childhood mortality, the difference between men and women at ages before 50 is an illustration of the impact of maternal mortality.

How representative are the data in a demographic database?

A demographic database based on Church records does not give a full account of the life history of all individuals contained in the database. Emigration and missing dates are the two main reasons for this. We will explore to what extent these factors pose a problem for the analyses and the interpretation of the results.

Moving out of the region: emigration

With the prospering of the industries in New England in the United States of America, an emigration wave occurred from the whole of rural Quebec (Lavoie, 1979). The Church records only contain registrations of children of emigrating married women from before the date of emigration and we do not know about eventual later born children or about the end of marriage. Therefore, we had to restrict our selection of subjects to married women who did not emigrate, i.e. their marriage ended (by death of themselves or their partner) in the SLSJ region (86% of the married women). In our analyses, we assumed that the women who stayed in the region did not form a selection with respect to fertility.

It may be that being childless made emigration easier or more attractive. On the other hand, it may just as well be that the economic necessity to emigrate was larger for couples with many children. These presumptions are hard to verify: the mere fact that the woman emigrated limits the amount of time she spent in the region to produce children. Therefore, emigrating women will have fewer children in the Church registers than women who stayed in the region and more of them will be without child. Indeed, the average number of registered births was 5.0 for emigrating women and 7.8 for women who stayed in the region and the proportion childlessness was 18% and 8% respectively. We cannot tell from these figures whether emigration was related to fertility. However, we can compare the time between marriage and birth of the first child, for women who had at least one child, assuming that less fertile women take longer time to become pregnant. The medians and 25^{th} to 75^{th} percentiles were 11.5 (9.8 – 15.5) months and 11.5 (9.8 – 15.2) months respectively for emigrating women

and for women who stayed in the region (P = 0.6, Mann-Whitney test). This tells us that the time to pregnancy of emigrating women who had a child before they left the region, was not different from women who stayed in the region. The emigrating women who did *not* have a child before they left the region, may have had a child (shortly) after that time. It then depends on whether they left the region soon after marriage or after a long period of time: in the former case they are likely to be as fertile as women who stayed in the region, whereas in the latter case they are likely less fertile, since their time to first pregnancy was longer than average. We conclude that: women who emigrated and had children at that time were as fertile as women with children who stayed in the region. Women who emigrated without children may have been less fertile, but this depends on how soon after marriage they emigrated.

Missing birth dates of children

For some women, the Church registers mention one or more children without a birth date. We cannot establish with certainty the ages of these women at first and last childbirth and they are therefore excluded from analyses. However, this exclusion biases the estimation of the probability to be childless, because the women with children are underrepresented. To deal with this problem we used 'sampling weights' that down-weigh the women without children. To give an example: in a group of N women, 4% remained childless and 96% had children, but in 10% of these women the birth date of one or more of their children was missing, and they are left out of the analysis. The number of women in the analysis is then: 0.04*N women without children, and (1-0.10)*0.96*N women with children. The estimate of childlessness is 0.04*N / (0.04*N + (1-0.10)*0.96*N) = 4.4%, instead of 4%. We compensate for this bias by assigning a sampling weight of (1-0.10) to the women without children. The corrected estimate of becomes (1-0.10)*0.04*N / ((1-0.10)*0.04*N + (1-0.10)*0.96*N) = 4%.

Two remarks should be made here:

1. We chose to down-weight the women without a child, where we might just as well have 'up-weighted' the women with children. However, we are going to use life-table methods to estimate childlessness, in which the women who remained childless are censored at the date of end of marriage. The statistical information in that type of analysis is solely determined by the number of women with children, and 'up-weighting' of these women would create more information than is present in the data, leading to too narrow confidence intervals.

2. We will also investigate differences in fertility within the group of women with children. The exclusion of women with missing childbirth dates should then be non-differential with respect to fertility. It seems unlikely that this requirement is fulfilled, for women with many children will have a higher chance that at least one of the birth dates is missing than women with few children. A logistic

regression analysis revealed that the chance of having children with missing birth dates indeed increased with higher number of children, from 7.4% for women with only one child to 12.7% for women with 20 children (P = 0.015). However, no such association was found in the subgroup of women who remained married until at least age 50, which is the group that will be used in these analyses (see below).

RESEARCH QUESTIONS

As said before, we will try to answer research questions from reproductive medicine in a historical natural population.

The first question, inspired by the observation that some women can get children up till late ages where others experience their end of fertility at a very young age is:

1. **What is the variation in female reproductive life span? Is the reproductive life span correlated to fertility at young age and if so, how strong is this correlation?**

Women in modern western societies are postponing reproduction to an age that is compatible with their lifestyle and personal circumstances. An issue of progressive concern is how long women can wait before their chances of becoming pregnant and having a child deteriorate. It has been argued that both the age-related decline and the end of female fertility are determined by the same biological process. If this holds, we expect that fertility at young age is positively related to the female reproductive life span. The aim of this study is first to evaluate in the BALSAC database the degree of variability of the reproductive female life span. Second, to evaluate whether fertility at young age is correlated to the reproductive life span and thereby explains some of this variation.

The second question is inspired by the clinical decision problem of when to start treatment in women who are diagnosed as infertile.

2. **Are fertility chances after 12 months trying to conceive small, as implied by the current standard of calling a couple infertile after 12 months?**

Several studies have published models for the prediction of spontaneous pregnancy chances among infertile patients, based on follow-up of actual patients. In all these studies, the follow-up was limited because many patients started treatment soon after inclusion. Instead, data on a historical population without treatment or contraception may be used to estimate the spontaneous pregnancy prognosis of infertile couples. The aim of this study is to assess the spontaneous pregnancy prognosis of infertile couples, depending on the woman's age, the duration of infertility and parity.

Selection of subjects

For our studies on natural fertility, the subjects should be married and their date of end of marriage should be known in the database, as stated above. In some cases we restricted the sample further to women whose marriage lasted at least until their 50th birthday. The selection is depicted in detail in **Figure 4**.

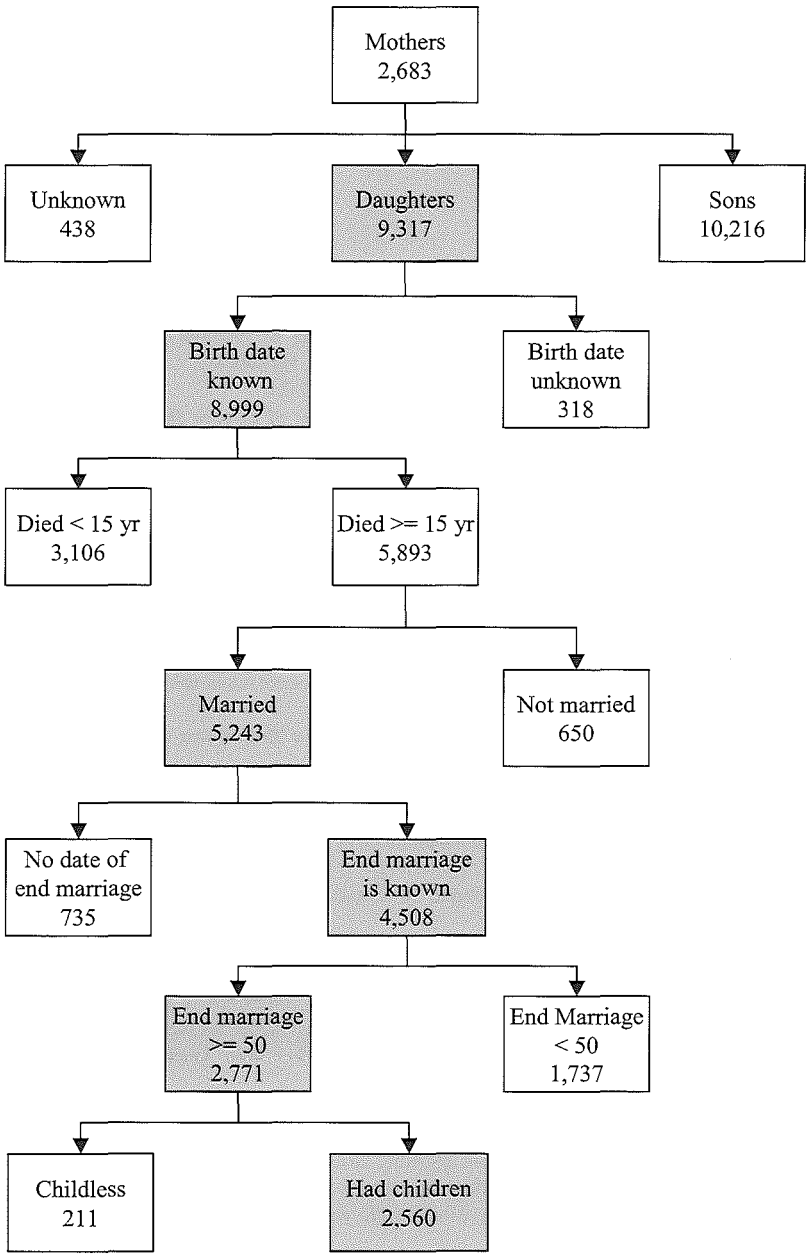


Figure 4 Selection of subjects from the SLSJ population.

OUTLINE OF PART I

In *chapter 2*, research question 1 is answered by analyzing the variability of age at last childbirth and the association between two measures of fertility at young age and age at last childbirth, on the BALSAC database.
In *chapter 3* we give an answer to research question 2 by analyzing the time from marriage to first childbirth and the time between first and second childbirth in the BALSAC database.
In *chapter 4*, the shortest chapter, the importance of the time-to-pregnancy concept is stressed, and its relation to the monthly pregnancy chances is explained.

References

Barlow DH. (2003) The design, publication and interpretation of research in Subfertility Medicine: uncomfortable issues and challenges to be faced. *Hum Reprod.* 18, 899-901.

Barrett JC. (1986) The estimation of natural sterility. *Genus.* 42, 23-31.

Bélanger C. The Quiet Revolution. *Events, Issues and Concepts of Quebec History.* Montreal, Quebec: Marianopolis College, 2000.

Blossfeld HP. (1995) *The new role of women: family formation in modern societies*, Westview Press,Boulder, Colorado.

Bouchard G, Roy R, Casgrain B, Hubert M. (1989) Fichier de population et structures de gestion de base de données: le fichier-réseau BALSAC et le système INGRES/INGRID. *Histoire & Mesure.* 4, 39-57.

Bouchard G, Roy R. (1991) Fécondité et alphabétisation au Saguenay et au Québec (XIX-XXe Siècles). *Annales de Démographie historique.* 173-201.

Bouchard G. (1996) *Quelques arpents d'Amerique. Population, économie, famille au Saguenay, 1838-1971.*, Boréal,Montréal.

Bouchard G. (2000) Through the meshes of patriarchy: the male/female relationship in the Saguenay peasant society (1860-1930). *The history of the family.* 4, 397-425.

Collins JA, Burrows EA, Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril.* 64, 22-8.

Dixon RB. (1971) Explaining cross-cultural variations in age at marriage and proportions never marrying. *Population Studies.* 25, 215-233.

Gini C. (1924) Prémières recherches sur la Fécondabilité de la femme. *International mathematical congress.* 2, 889-892.

Henripin J. Tendances et facteurs de la fécondité au Canada. Ottawa: Bureau Fédéral de la Statistique, 1968:30.

Henry L. (1961) Some data on natural fertility. *Eugenics quarterly.* 8, 81-91.

Henry L. (1976) *Population: analysis and models*, Edward Arnold,London.

Lalou R. La haute fécondité des Québécoises avant la révolution contraceptive: effet de nature ou effet de culture? Un bilan des connaissances. Chicoutimi: Université du Québec à Chicoutimi, 1990.

Lavoie Y. L'Émigration des Québécois aux États-Unis de 1840 à 1930. Québec: Conseil de la langue française, 1979.

Leridon H. (1998) [30 years of contraception in France] 30 ans de contraception en France. *Contracept Fertil Sex.* 26, 435-8.

Menken J, Trussell J, Larsen U. (1986) Age and infertility. *Science.* 233, 1389-94.

Potter RG, Parker MP. (1964) Predicting the time required to conceive. *Population studies.* 18, 99-116.

Schwartz D. (1981) Importance de la durée d'infécondité dans l'appréciation de la fertilité d'un couple. *Population.* 2, 237-250.

Schwartz D, Mayaux MJ. (1982) Female fecundity as a function of age: results of artificial insemination in 2193 nulliparous women with azoospermic husbands. Federation CECOS. *N Engl J Med.* 306, 404-6.

Snick HK, Snick TS, Evers JL, Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod.* 12, 1582-8.

te Velde ER. Zwanger worden in de 21-ste eeuw: steeds later, steeds kunstmatiger. Rede uitgesproken bij de aanvaarding van het ambt van bijzonder hoogleraar in de Leer van de Menselijke Vruchtbaarheid aan de Rijksuniversiteit te Utrecht. Utrecht: Rijksuniversiteit Utrecht, 1991.

Trussell J, Wilson C. (1985) Sterility in a population with natural fertility. *Population studies.* 39, 269-286.

Vail A, Gardener E. (2003) Common statistical errors in the design and analysis of subfertility trials. *Hum Reprod.* 18, 1000-4.

van Noord-Zaadstra BM, Looman CW, Alsbach H, Habbema JD, te Velde ER, Karbaat J. (1991) Delaying childbearing: effect of age on fecundity and outcome of pregnancy. *Bmj.* 302, 1361-5.

Von Cube A. (1986) Europeans: an endangered species? *Popul Today.* 14, 6-7.

Wood JW. (1994) *Dynamics of human reproduction: biology, biometry, demography*, Aldine de Gruyter,New York.

2

The end of female fertility is highly variable and related to fertility at young age

ABSTRACT

Introduction: Women in modern western societies are postponing reproduction. An issue of progressive concern is until which age women can still have a child. The first aim of this study was therefore to evaluate the degree of variability in female reproductive life span. The second aim, motivated by the theory of ovarian ageing, was to evaluate whether the reproductive life span is correlated to fertility at young age. We investigated these research questions in a group of women from a 19th century natural fertility population in Quebec, Canada.

Materials and methods: We included 2,488 women born between 1840 and 1899, whose marriage lasted at least until age 50. The Kaplan-Meier method on age at last childbirth was used to determine the age-specific chance of being effectively infertile, i.e. no further become pregnant leading to live birth. We used two measures of fertility at young age: the birth rate between ages 18 and 30 and the maximum number of children born within a five-year period. Spearman's correlation coefficients of the measures of fertility at young age with age at last childbirth were determined and corrected for the potentially confounding variables: year of birth, age at marriage and occupation of the husband as an indicator for socio-economic status.

Results: The chance that a woman no longer conceives a pregnancy leading to live birth was 10% at age 29 years. It increased to 25%, 40%, 50%, 60%, 75% and 90% at ages 37, 39, 40.5, 41, 42.5 and 44 years respectively. The birth rate between ages 18 and 30 and the maximum number of children born within a five-year period were positively correlated with age at last childbirth ($r = 0.15$ and 0.14 respectively, $P < 0.001$). The correlation was absent in women of above average fertility, but stronger for women of below average fertility ($r = 0.22$ and 0.28 respectively). Correction for confounders had no major influence on results.

Conclusions: The end of effective fertility shows a large variation. There is evidence that part of this variation is explained by fertility at young age.

INTRODUCTION

Since the introduction of effective birth control at the end of the sixties, extraordinary changes in reproductive behaviour have taken place, such as voluntary childlessness, rapid decrease in the number of children per couple and delays in childbearing (van de Kaa, 1987). Having children became no longer the unavoidable destiny of a woman, but an issue for careful consideration and planning. Many couples nowadays delay childbearing to a period in life that is compatible with their lifestyles and personal circumstances.

However, an issue of progressive concern is how long women can wait before their chances of having a child deteriorate. Because of the wide-scale use of contraceptive measures in modern western populations, this problem can only be studied in so-called natural fertility populations in which birth control was not or hardly practiced and religious convictions urged couples to produce as many children as long as possible. Under such circumstances, age at last childbirth can be considered as the best proxy for the age at which effective female fertility, i.e. the ability of a woman to conceive a pregnancy that will end in a live birth, comes to an end (Bongaarts, 1982). Studies in natural fertility populations have shown that fertility declines with age in a manner that is universal throughout the human species (Leridon, 1977) (Bongaarts, 1982) (Spira, 1988) (Wood, 1989). Although decreasing coital frequency also plays part, the female age-related contribution to this decline appears to be of paramount importance (Weinstein *et al.*, 1993).

It has been argued that the decline -from birth onwards- of both the quantity and the quality of the oocyte/follicle pool determines the age–dependent loss of female fertility (Faddy *et al.*, 1992) (te Velde *et al.*, 1998). If this concept holds, we expect that women with a slow rate of decline have a higher fertility at young age and also a later onset of sterility compared to women with a faster decline. As a consequence, fertility at young age would be positively related to the onset of female sterility. The aim of this study was first to evaluate the degree of variability in reproductive female life span. Second, to evaluate whether fertility at young age is correlated with the reproductive life span.

MATERIALS AND METHODS

Subjects

The study used data from women born between 1840 and 1899 in the Saguenay-Lac-St-Jean (SLSJ) region in Quebec Canada, who married and died in the region. The SLSJ region became populated by colonists from the nearby Charlevoix region around the middle of the 19th century and their data from church registers (Roman-Catholic) on birth, baptism, marriage and death are contained in the BALSAC demographic

database (Bouchard *et al.*, 1989). Due to their tradition of short lactation, the total number of children per woman was high in comparison with other natural populations (Lalou, 1990), which makes them ideally suited to study human fertility under natural conditions. The demographic fertility transition occurred relatively late in this region, in 1946-1955 among farmers and in 1931-1935 for all other professions (Bouchard and Roy, 1991). For this reason, the birth cohorts chosen were restricted to those between 1840 to 1899, to limit the effect of the use of contraceptives on the apparent fertility of the women.

In natural populations, age at last childbirth is a marker of the end of effective fertility when a woman remains married long enough. Therefore, only women whose marriage lasted at least until the age of 50 were considered (in demography, this is the age at which a woman's reproductive career is considered to be completed). We assumed that marriages ended with the death of the woman or of her husband, since divorce was extremely rare in this Church-dominated population (Bouchard, 2000).

Fertility at young age

Evidence from artificial insemination programs in which the women did not have an apparent impairment of fertility showed that pregnancy chances are optimal from the age of 18 years onwards, and start to decline only after the age of 30 (Schwartz and Mayaux, 1982) (van Noord-Zaadstra *et al.*, 1991). This motivated us to measure fertility at young age for an individual woman by her birth rate during the age period 18 to 30. We defined the birth rate as the number of childbirths between ages 18 and 30, divided by the number of married years in that age interval. For women who married after the age of 18, only the years between marriage and the 30th birthday contributed to the married years. A woman should have at least one childbirth between ages 18 and 30. This measure of individual fertility is similar to the age-specific fertility rate, a population measure commonly used by demographers (Newell, 1988).

As a second measure of fertility at young age we used the maximal number of births in five successive years (NMAX) (Le Bourg *et al.*, 1993). This parameter measures the peak reproductive performance of a woman, which not necessarily occurs at young age for all women, but will do so for most of them.

Data analysis

The age at which effective fertility ended was estimated using both the age at last childbirth for the women who had children and the age at marriage of the women who remained childless. These latter women might have had a childbirth (and consequently also a last childbirth), if they had married at a younger age. Therefore, we included them in the analysis by assuming that their chances of having a last childbirth at the ages before their marriage were equal to the chances of the women who were married at those ages. Technically, this analysis was performed using the Kaplan-Meier method, with time running backwards from age 55 onwards (the last age at which a woman gave birth to a child was 54). The (55 minus) age at last childbirth was the

time of event and (55 minus) age at marriage was the censoring time for the women who remained childless. The resulting Kaplan-Meier estimates were plotted against the original age-axis (subtracting 9 months to account for the gestational period between conception and live birth) to give the curve of the age at which effective fertility came to an end. We tested whether the results depended on the age at which the woman married, as has been suggested in the literature (Trussell and Wilson, 1985), by using the log-rank statistic on age at marriage in 5-year groups.

There were women who had one or more children mentioned in the database, without their birth date. Consequently, we didn't know which child was the lastborn and so these women had to be excluded (11% of the women with children). Because of this exclusion, women with a child are under-represented in the analysis, which made a correction of the Kaplan-Meier estimates afterwards necessary.

The associations between the two measures of fertility at young age and age at last childbirth were quantified by Spearman's rank correlation. In order to have at least 5 married years available between ages 18 and 30, the data were restricted to women who were married before age 25. Further, we excluded women who had their last childbirth before the age of 30, because in these women the birth rate between ages 18 and 30 would (by definition) be reduced and therefore correlated 'by construction' with age at last childbirth.

Literature shows that the occupation of the husband, the woman's year of birth (Bouchard and Roy, 1991) (Lycett *et al.*, 2000) and her age at marriage (Trussell and Wilson, 1985) are associated both with fertility at young age and age at last childbirth. Therefore, we considered these factors as potential confounders, and we corrected our analysis using the following categories: occupation of the spouse in three classes ('farmer', 'blue collar' and 'white collar'), year of birth from before 1860 and until 1890-1899 in five 10-year intervals, and age at marriage in five classes ('under 17', four two-year intervals from 17 to 25). To quantify the strength of association between the measures of fertility at young age and age at last childbirth after adjustment for confounders, Spearman's rank correlation was calculated between the measures of fertility at young age and the residuals from an ANOVA for age at last childbirth on the confounding factors.

The parameters of fertility measure the overall effect of biological potential and behaviour. Particularly breastfeeding may be an important behavioural factor determining apparent fertility, by lengthening the interval between births. We estimated the period of breastfeeding for each woman by taking the difference between the interval from first to second birth and the interval from marriage to first birth (Sheps, 1965).

A p value below 0.05 was considered to denote statistical significance.

RESULTS

Initially, we selected 4,508 married women born between 1840 and 1899 from the database, but 1,737 of them had to be excluded because their marriage ended before their 50th birthday. Of the selected 2,771 women, 211 (7.6%) remained childless. For 283 of the 2,560 women with child (11%), the age at last childbirth was unknown, which excluded them from analysis, but which was corrected for afterwards. The analysis determining the age at which the effective fertility ended was therefore based on 2,488 women (= 2,277 with child + 211 childless).

In this selection, the median age at last childbirth was 41.4 years, with a range from 17 to 54 years. The Kaplan-Meier curve for the age at which the effective fertility ended is shown in **Figure 1**. The curve shows a moderate increase that starts to accelerate at age 30 and becomes more profound from age 35 onwards. The chance that a woman no longer conceives a pregnancy leading to live birth was 10% at age 29 years. It increased to 25%, 40%, 50%, 60%, 75% and 90% at ages 37, 39, 40.5, 41, 42.5 and 44 years respectively. The Kaplan-Meier curve did not depend on the age at which the woman married (P = 0.2).

The second analysis, about the relationship between fertility at young age and age at last childbirth required a further selection of women: of the previously selected 2,771 women (see above), 2,239 were married before age 25, 109 of whom were childless

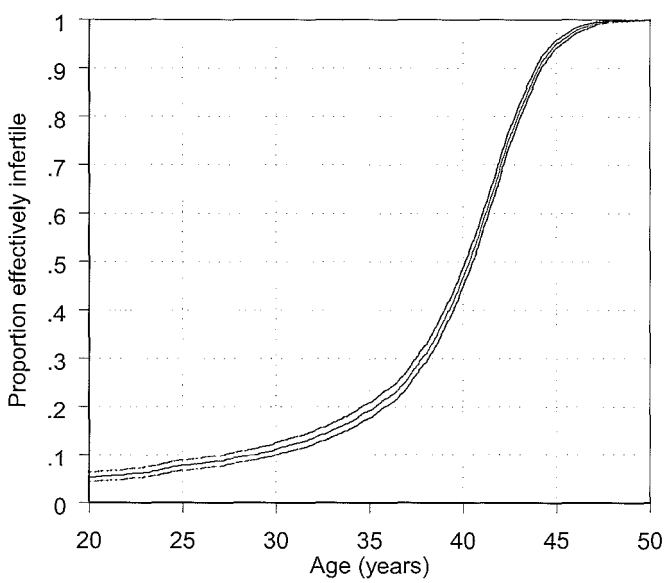


Figure 1 Cumulative age at the end of effective fertility. Kaplan-Meier estimates for 2,488 women, with 95% confidence interval.

Table 1 Age at last childbirth, total number of births, birth rate between age 18 and 30, number of births, five-year interval of maximal fertility and estimated duration of breastfeeding for 1,755 women from the SLSJ region (Quebec, Canada).

(a) Descripive statistics

	Mean	(Std. Dev)	Median	(P10-P90)*
Age at last childbirth (years)	41.0	(3.6)	41.6	(35.9-45.0)
Total number of births	11.0	(3.1)	11	(7-15)
Birth rate between age 18 and 30 (per year)	0.58	(0.14)	0.58	(0.42-0.75)
Maximal number of births in five successive years	3.7	(0.7)	4	(3-5)
Five year interval of maximal fertility (midpoint)	25.4	(4.2)	24.5	(20.9-31.6)
Estimated period of breastfeeding (months)	6.8	(13.6)	6.3	(-3.0-15.2)

*P10 and P90 represent the 10th and 90th percentiles.

(b) Spearman's rank correlation coefficients

	ALAS	BR 18-30	NMAX
Age at last childbirth (ALAS)	1	.15	.14
Birth rate between age 18 and 30 (BR 18-30)	.15	1	.66
Maximal number of births in five successive years (NMAX)	.14	.66	1
Estimated period of breastfeeding	-.07	-.21	-.27

Note: Correlations coefficients larger in size than 0.05 are statistically significant

and in 252 women the age at last childbirth was unknown. Of the remaining 1,878 women, 109 had their last childbirth before age 30 and another 14 women did not have a childbirth before the age of 30. Since we wanted to correlate fertility at young age with the end of effective fertility these women (109 + 252 + 109 + 14) had to be deleted, and therefore this analysis is based on 1,755 women.

The characteristics of these women are shown in **Table 1**(a), and their pairwise correlations in **Table 1**(b). Both the birth rate between ages 18 and 30 and the maximum number of births within a five year period showed a positive association with age at last childbirth ($r = .15$ and $r = .14$ respectively, both $P < 0.001$). The correlation with the estimated duration of breastfeeding was negative, as to be expected ($r = -.21$ and $-.27$ respectively).

The nature of the association between age at last childbirth and the birth rate between ages 18 and 30 is illustrated in **Figure 2**. A steady increase in age at last

childbirth with increasing birth rate per year is seen, from 37 years at a birth rate of 0.2 to a maximum of 41 years for birth rates of 0.5 and higher. The distribution of age at last childbirth, depicted on the vertical axis, is skewed, with some women having their last childbirth already at a very young age. The distribution of the birth rate between 18 and 30 is shown on the horizontal axis and is more symmetrical, with most women

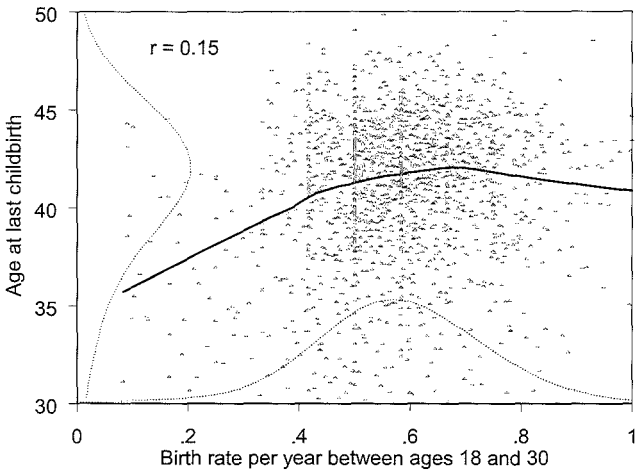


Figure 2 Correlation between age at last childbirth and birth rate between ages 18 and 30. On the axis, the distribution of the data is depicted. N = 1,755.

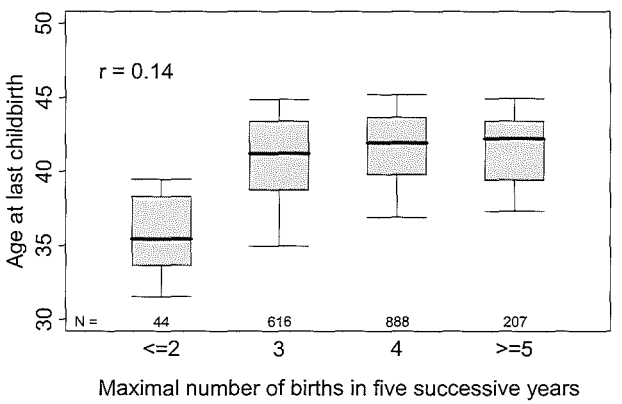


Figure 3 Correlation between age at last childbirth and the maximal number of births in five successive years. Box and whisker plot, with boxes representing median and 25th and 75th percentiles and whiskers representing 10th and 90th percentiles. N = 1,755.

having a birth rate between 0.4 and 0.8 births per year.

The association between the maximal number of births in five successive years and age at last childbirth is depicted in **Figure 3**. The same levelling off in trend is observed as in **Figure 2**. The trend is present until a maximum of 4 children per 5 years. For the highest group (>= 5 births) a minor further increase in age at last childbirth is seen, whereas in Figure 2 the curve seems to go down a little.

The distribution of the confounding factors and their relationship with age at last childbirth and fertility at young age are shown in **Table 2**. Occupation of the husband was significantly associated with age at last childbirth, which was one year later for the wives of farmers compared to the 'blue collar' and 'white collar' groups. No significant association was present with fertility at young age. A more recent year of birth was significantly associated with earlier age at last childbirth, but at the same time with an elevated level of the maximal number of births in five years. Finally, younger age at marriage was associated with a lower birth rate between ages 18 and 30, with a slightly higher maximal number of births in five years, but not with age at last childbirth. Adjustment for the confounding factors hardly changed the association between age at last childbirth and the measures of fertility at young age (adjusted r = 0.15 and 0.15 for the birth rate between ages 18 and 30 and the maximum number of birth in five years respectively, both $P < 0.001$).

DISCUSSION

Female fertility and the age at which it comes to an end can best be studied in so-called natural fertility populations, in which birth control is not (or hardly) practiced (Menken *et al.*, 1986). The mean age for the birth of the last child are surprisingly similar in most natural fertility populations all over the world, including contemporary ones (Wood, 1994). Age at last childbirth can be regarded as the age when the effective fertile period of a woman comes to an end. Many demographers regard the theoretical age at sterility as the end of fertility (the capacity of a woman to conceive irrespective of the result of pregnancy) (Leridon, 1977) (Trussell and Wilson, 1985) (Wilson *et al.*, 1988) (Larsen and Menken, 1989). However, for the aim of counseling women about how long they can postpone childbearing, age at last child birth is of greater importance than the age at sterility, because couples want to have an estimate of the period of time they still can have children.

Our data demonstrate that the end of the effective fertile period of women is extremely variable, with a median age of 41.4 years ranging from 17 to 54 years. In most women, the effective fertile period ends between 35 and 45 years. But still in about 19% of the women this period ends before 35 years and in 5% it ends after age 45 (**Figure 1**). A similar study has been performed by Trussell and Wilson using the sterility definition

Table 2 Age at last childbirth, birth rate between ages 18 and 30 and maximal number of births in five successive years, by the husband's occupation, year of birth and age at marriage, for 1,755 women from the SLSJ region (Quebec, Canada).

	Number	Age (yr) at last childbirth Mean (SD)	Birth rate/year ages 18 - 30 Mean (SD)	Maximal number of births in 5 years Mean (SD)
Husband's occupation	*P-value*[a]	*< 0.001*	*0.07*	*0.6*
Farmer	1098	41.4 (3.5)	.59 (.13)	3.7 (.7)
Blue collar	562	40.4 (3.7)	.57 (.14)	3.7 (.7)
White collar	77	40.4 (4.0)	.57 (.17)	3.7 (.9)
Year of birth	*P-value*[b]	*< 0.001*	*0.2*	*< 0.001*
<1860	168	41.9 (3.4)	.57 (.13)	3.6 (.7)
1860-69	296	41.4 (3.1)	.58 (.12)	3.7 (.7)
1870-79	487	41.0 (3.4)	.58 (.14)	3.7 (.7)
1880-89	567	40.9 (3.8)	.58 (.14)	3.8 (.7)
1890-99	237	40.1 (4.2)	.58 (.16)	3.8 (.8)
Age at marriage	*P-value*[b]	*0.8*	*< 0.001*	*0.02*
<=16	138	41.1 (3.4)	.54 (.12)	3.9 (.7)
17-18	428	41.1 (3.6)	.55 (.12)	3.7 (.7)
19-20	562	41.0 (3.6)	.58 (.13)	3.7 (.7)
21-22	370	40.8 (3.6)	.60 (.14)	3.7 (.8)
23-24	257	40.9 (3.8)	.61 (.16)	3.6 (.7)
Total	1755	41.0 (3.6)	.58 (.14)	3.7 (.7)

[a, b] P-values from analysis of variance(ANOVA), F-test[a] or test for trend[b].

(Trussell and Wilson, 1985). For ages below 35, their estimates were remarkably similar to ours. However, at higher ages their estimates of sterility chances were considerably lower, after they corrected their data for age at marriage. Since older female age is associated with an increased risk of pregnancy loss, it makes sense that our estimates of effective infertility are higher than the demographic estimates of sterility. A remarkable finding in the report by Trussell and Wilson was that women who married at a young age appeared to have an earlier onset of permanent sterility. They speculate that this phenomenon is caused by infections, occurring during or after childbirth, that resulted in sterility due to tubal disease. A consequence of early marriage is then that a woman is exposed to these risks from an early age onwards. We could not see such an effect in our data: earlier age at marriage was not related with a reduction of effective fertility. The data most commonly used by demographers are usually older than ours -mostly pre-industrial (Menken *et al.*, 1986) (Wood, 1994)- and it may well be that the risk of infections during childbirth was higher in those times than in our population living at the end of the 19th and beginning of the 20th

century, where the knowledge of modern health care practices, such as better hygiene, were possibly already (partly) implemented (Loudon, 1986). Therefore, we regard our data as more representative for modern times than the pre-industrial populations used by Trussell and Wilson.

We excluded many women because their marriage ended before their 50th birthday, since in these women age at last childbirth cannot be observed. In 65% of cases, marriages ended before age 50 because the woman had died. One may wonder whether the selected women did not form a healthy sub-group that is not representative in fertility for the total population. However, the two measures of fertility at young age in women who were excluded were very similar to the values in the selected women: the birth rate between ages 18 and 30 was on average 0.58 (SD = 0.13) and the maximal number of children in five successive years 3.7 (SD = 0.8), comparable with the values in Table 1. Therefore, we conclude that the selection of women that we used in our analysis was representative for the population.

What does this "lesson from the past" teach us? First, in the present era of postponing the time to have children, women should be aware of these facts. Unfortunately, there are no early signs of premature reproductive failure. In a previous study we demonstrated (van Zonneveld *et al.*, 2003) that before the period of irregular menstrual cycles and climacteric complaints, there is a long period of infertility during which women still have normal ovulatory cycles and normal hormone levels. This period is likely to be preceded by many years of progressively declining pregnancy chances. Even the most advanced assisted reproductive technique comes too late in such a situation and the only way still to become pregnant and have a child -allbeit genetically different- is to resort to oocyte donation. Second, that the end of the female reproductive period is extremely variable. The chances to have the end of the effective fertile period before age 35 are relatively low, about 17%. Would that be a too pessimistic estimate for the present time? In a previous study we compared the age distributions of age at menopause in a contemporary population and age at last childbirth in the SLSJ population (te Velde and Pearson, 2002). The almost equal distributions of both populations, including the skewness to younger ages, not only suggest that mean ages differ by about ten years, but that individual women also experience a ten year difference between the age of becoming infertile and age at menopause. The striking similarity not only suggests that premature ovarian failure probably did occur in this 19th century population, but also that the end of female fertility and age at menopause are part of the same process.

Recently, we have summarized other arguments suggesting that the age-dependant loss of female fertility is dictated by the decline of both the quantity and quality of the oocyte pool (te Velde and Pearson, 2002). During early fetal life the ovaries are packed with the entire stock of follicles (oocytes surrounded by ovarian granulose cells), but already before birth, numbers of follicles decline in an exponential fashion.

From the millions present before birth, some 100.000 are left at the beginning of puberty and when the menopause is reached at a mean age of 51 years, the supply is reduced to less than thousand, a number insufficient to sustain the hormonal changes necessary for menstruation (Faddy *et al.*, 1992). The essence of this so-called ovarian concept is that after a period of optimal fertility from age 18-30 years, oocyte quality decreases in parallel to the progressive loss of follicle numbers and that menopause is the final reproductive event, marking the definite exhaustion of the follicle pool. It is the only event in the cascade of changes involved in the ageing process that can be noticed unambiguously by individual women. Determination of the ages and periods of the preceding events, such as the beginning of subfertility and infertility is much more difficult to establish. The mean age of menopause is 50-51 years in all western countries, with a huge variation from 40-60 years (Treloar, 1981). According to the ovarian concept, we would expect that fertility at young age would positively correlate with the end of the fertile period. Indeed, we found a positive relation between fertility at young age and age at last childbirth as a proxy for the end of effective female fertility, confirming this concept.

We used two different measures of fertility at young age, the birth rate between ages 18 and 30 and the maximal number of children within a five-year period. While the first parameter measures –by definition- fertility at young age, the second one measures the peak fertility of a woman (Le Bourg *et al.*, 1993), which not necessarily occurs at young age for all women, but usually does: the two measures showed a positive correlation (r = 0.66, **Table 1b**). The relation between the first measure and age at last childbirth was obvious until a fertility level of about 1 birth in every two years between ages 18 and 30. Given that the period of breast-feeding was estimated to be on average six months (**Table 1a**), this probably is the maximum achievable in highly fertile women. Because we found that higher fertility at young age was associated with shorter periods of breastfeeding (see **Table 1b**), we conclude that the women who showed higher levels of fertility at young age probably accomplished this by breastfeeding their children for a shorter period. The second measure, the maximal number of births during 5 years, confirmed this pattern (**Figure 3**). In the subgroup of women who have a birth rate of less than 0.6 per year, the correlation coefficients with age at last childbirth became 0.22 and 0.27 instead of the correlations of 0.15 and 0.14 presented in **Table 1(b)**. Given the fact that many other variables including coital frequency and fertility of the male partner, but also unknown factors originating from life style and environment, must have influenced the age at which the female reproductive period ended, we consider these correlations as surprisingly high.

In summary, we conclude that the end of effective female fertility is extremely variable and that part of this variation is explained by fertility at young age.

References

Bongaarts J. The proximate determinants of natural marital fertility. New York: Population Council: Center for Policy Studies, 1982:1-43.

Bouchard G, Roy R, Casgrain B, Hubert M. (1989) Fichier de population et structures de gestion de base de données: le fichier-réseau BALSAC et le système INGRES/INGRID. *Histoire & Mesure.* 4, 39-57.

Bouchard G, Roy R. (1991) Fécondité et alphabétisation au Saguenay et au Québec (XIX-XXe Siècles). *Annales de Démographie historique.* 173-201.

Bouchard G. (2000) Through the meshes of patriarchy: the male/female relationship in the Saguenay peasant society (1860-1930). *The history of the family.* 4, 397-425.

Faddy MJ, Gosden RG, Gougeon A, Richardson SJ, Nelson JF. (1992) Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. *Hum Reprod.* 7, 1342-6.

Lalou R. La haute fécondité des Québécoises avant la révolution contraceptive: effet de nature ou effet de culture? Un bilan des connaissances. Chicoutimi: Université du Québec à Chicoutimi, 1990.

Larsen U, Menken J. (1989) Measuring sterility from incomplete birth histories. *Demography.* 26, 185-201.

Le Bourg E, Thon B, Legare J, Desjardins B, Charbonneau H. (1993) Reproductive life of French-Canadians in the 17-18th centuries: a search for a trade-off between early fecundity and longevity. *Exp Gerontol.* 28, 217-32.

Leridon H. (1977) *Human fertility: the basic component*, Chicago University Press,Chicago.

Loudon I. (1986) Deaths in childbed from the eighteenth century to 1935. *Med Hist.* 30, 1-41.

Lycett JE, Dunbar RI, Voland E. (2000) Longevity and the costs of reproduction in a historical human population. *Proc R Soc Lond B Biol Sci.* 267, 31-5.

Menken J, Trussell J, Larsen U. (1986) Age and infertility. *Science.* 233, 1389-94.

Newell C. (1988) *Methods and models in demography*, Guilford Press,New York.

Schwartz D, Mayaux MJ. (1982) Female fecundity as a function of age: results of artificial insemination in 2193 nulliparous women with azoospermic husbands. Federation CECOS. *N Engl J Med.* 306, 404-6.

Sheps MC. (1965) An analysis of reproductive patterns in an american isolate. *Population studies.* 19, 65-80.

Spira A. (1988) The decline of fecundity with age. *Maturitas.* Suppl 1, 15-22.

te Velde ER, Dorland M, Broekmans FJ. (1998) Age at menopause as a marker of reproductive ageing. *Maturitas.* 30, 119-25.

te Velde ER, Pearson PL. (2002) The variability of female reproductive ageing. *Hum Reprod Update.* 8, 141-54.

Treloar AE. (1981) Menstrual cyclicity and the pre-menopause. *Maturitas.* 3, 249-64.

Trussell J, Wilson C. (1985) Sterility in a population with natural fertility. *Population studies.* 39, 269-286.

van de Kaa DJ. (1987) Europe's second demographic transition. *Pop. Bull.*, 42, 1-57.

van Noord-Zaadstra BM, Looman CW, Alsbach H, Habbema JD, te Velde ER, Karbaat J. (1991) Delaying childbearing: effect of age on fecundity and outcome of pregnancy. *Bmj.* 302, 1361-5.

van Zonneveld P, Scheffer GJ, Broekmans FJ, Blankenstein MA, de Jong FH, Looman CW, et al. (2003) Do cycle disturbances explain the age-related decline of female fertility? Cycle characteristics of women aged over 40 years compared with a reference population of young women. *Hum Reprod.* 18, 495-501.

Weinstein M, Wood J, Greenfield DD. (1993) How does variation in fetal loss affect the distribution of waiting times to conception? *Soc Biol.* 40, 106-30.

Wilson C, Oeppen J, Pardoe M. (1988) What is natural fertility? The modelling of a concept. *Popul Index.* 54, 4-20.

Wood J. Fecundity and natural fertility in humans. In: Milligen S, editor. *Reviews of Reproductive Biology.* Oxford: Oxford University Press, 1989:61-109.

Wood JW. (1994) *Dynamics of human reproduction: biology, biometry, demography*, Aldine de Gruyter,New York.

3

Primary and secondary infertility before the era of birth control: lessons from the past

ABSTRACT

Objective: To assess the spontaneous pregnancy prognosis of infertile couples, depending on the woman's age, the duration of infertility and parity.
Design: Analysis of demographic data.
Setting: The Saguenay-Lac-St-Jean region in Quebec Canada.
Patients: 4,125 married women, born between 1844 and 1900.
Interventions: Women who did not conceive their first child within 1 to 4 years after marriage or their second child within 1 to 4 years after the first childbirth were labelled as primary or secondary infertile. Age, duration and parity-dependent prognosis was estimated using Cox regression.
Main outcome measure: Pregnancy resulting in live birth.
Results: Of the women with 'primary' infertility of one year duration, 44% were predicted to conceive a child within the next year in the age group 20-24, and 39%, 32% and 22% respectively in the age groups 25-29, 30-34 and 35-39. At 2 years durations the age specific predictions were 26%, 21%, 17% and 12%. Predictions for 'secondary' infertility at one year duration were 63%, 56%, 48% and 36%, and at 2 years duration 38%, 33%, 28% and 19% respectively in the age groups 20-24, 25-29, 30-34 and 35-39.
Conclusions: In women under 35, after one year of primary infertility and 1-2 years of secondary infertility, prospects are good. If nothing is found diagnostically, prognosis may be better and therapy could be withheld for the next 1 or 2 years.

INTRODUCTION

In this era of assisted reproductive technology (ART), there is little attention for the chances of a couple to conceive spontaneously. Yet, in order to judge whether treatment is indicated, such knowledge is mandatory.

Studies on prognostic factors in patients with subfertility have shown that increased female age, longer duration of infertility and primary infertility have an adverse effect on the chance of spontaneous conception leading to live birth (Collins *et al.*, 1995 Eimers *et al.*, 1994 Snick *et al.*, 1997). In these studies, couples with a fertility problem were followed until pregnancy or start of treatment. The estimates of the effects of the prognostic factors on the probability of pregnancy appear to be consistent between these studies. However, since patients who got treated may be a selected part of the total group of couples with a fertility problem, the absolute level of the cumulative pregnancy curve ("how many couples will eventually become pregnant spontaneously?") and its shape ("how fast do they become pregnant?") may be biased. Only long-term follow-up of untreated couples can give truly reliable information on the chance to conceive spontaneously. In modern western populations such a study is unlikely to be performed, because treatment nowadays cannot be withheld for long. Data from populations where there is no evidence of birth control (which we will call natural populations), either non-western or historical, do not suffer from this problem, and therefore may reveal useful information for clinical practice.

We studied the spontaneous pregnancy prognosis for primary and secondary infertility, dependent on age and duration of infertility, in a historical natural population. The aim of our study was to present data on the chance to become pregnant spontaneously from a population without contraception or fertility treatment.

MATERIALS AND METHODS

Subjects and definitions

We selected 4,508 women, born between 1844 and 1900 in the Saguenay-Lac-St-Jean region in Quebec Canada, who had a registration of the date of marriage and the date of death of herself or of her husband. The Saguenay-Lac-St-Jean region became populated by immigrants from the nearby Charlevoix region around the middle of the 19th century and the data from church registers on birth, baptism, marriage and death are contained in the BALSAC demographic database (Bouchard *et al.*, 1989). We used the data concerning year and age at marriage, and the dates of birth of the first and second child. The birth cohort 1844 to 1899 was chosen to exclude as much as possible any disturbing effect of the use of contraceptive measures (Lalou, 1990). Of these 4,508 married women, 364 (8.1%) remained childless.

The following definitions were used: Infertility was used in the context of not having a child within a certain period. Primary infertility referred to not having a first child within a certain period following marriage, secondary infertility referred to not

having a second child within a certain period following the first childbirth. Childlessness denoted that a couple never had a child, but in practice the proportion childlessness was estimated at 10 years after marriage. We avoided the word 'sterility' since it suggests that childlessness was involuntary, which we can never know for certain in any population.

Missing birth dates

For 383 women, one or more children were registered in the database without their birth date. Consequently, we don't know for sure which child is the firstborn. We decided to exclude these women, leaving 4,125 for analysis. However, because of this exclusion the women with a child are under-represented in the analysis, which made a correction afterwards necessary. Missing birth dates of children occurred just as often in women who had a second child as in women who had only one child. Therefore the analysis of the time from first to second childbirth was based on the women without missing birth dates of their children, and no correction was needed here.

Data analysis

Kaplan-Meier analysis, stratified by age at marriage, was used to estimate the cumulative chance of first childbirth. The time between the date of marriage and the birth of the first child was the time variable. In case no first childbirth was registered, the observation was censored: the time variable was set equal to the end of marriage (defined by either death of the woman or of her husband, whichever occurred first). In a similar fashion the proportion of women who had a first child, but who remained without a second child was estimated with starting time set at date of first childbirth, and stratification by age at first childbirth. Cox regression with a flexible spline function was used to obtain a smoothed estimate of childlessness by age at marriage.

In order to extrapolate the findings of the present study to modern times, we made the assumption that a modern woman who starts trying to become pregnant for the first time at a certain age is comparable to a 19th century woman who married at that age. A fictitious diagnosis of primary infertility was established for women who did not have a first child conceived at 1, 2, 3 and 4 years after marriage (i.e. no child birth at 1, 2, 3 and 4 years after marriage, with a 9 months average gestational period added). The same was done for secondary infertility, starting from the time of first childbirth. In that case, the 'diagnosis' was established at 1 – 4 years duration since the first childbirth, with 9 months added for gestation. The average length of the period of post-partum amenorrhea due to breastfeeding was also added; it was estimated as the average difference between the time from first to second child and the time from marriage to the first child, for the women who had a second childbirth; the result was 6 months.

The data of these women formed a cohort that was used in a Cox regression analysis to estimate the age, duration of infertility and parity specific prognosis from the time of diagnosis of infertility, as it would have been established in a modern clinical

setting. The results were presented as fecundability ratios for five-year categories of age and 1-year categories of duration of infertility. The fecundability ratio (or hazard ratio in survival analysis) has the interpretation of relative risk of becoming pregnant (per cycle) in a certain category compared to the risk in a reference category.

Since the same woman could be included several times in this analysis (e.g. if she conceived her first child at 3.5 years after marriage she was included 3 times: with duration of infertility 1, 2 and 3 years), the data are not statistically independent. Therefore, the confidence intervals were corrected using Huber's method (Huber, 1967).

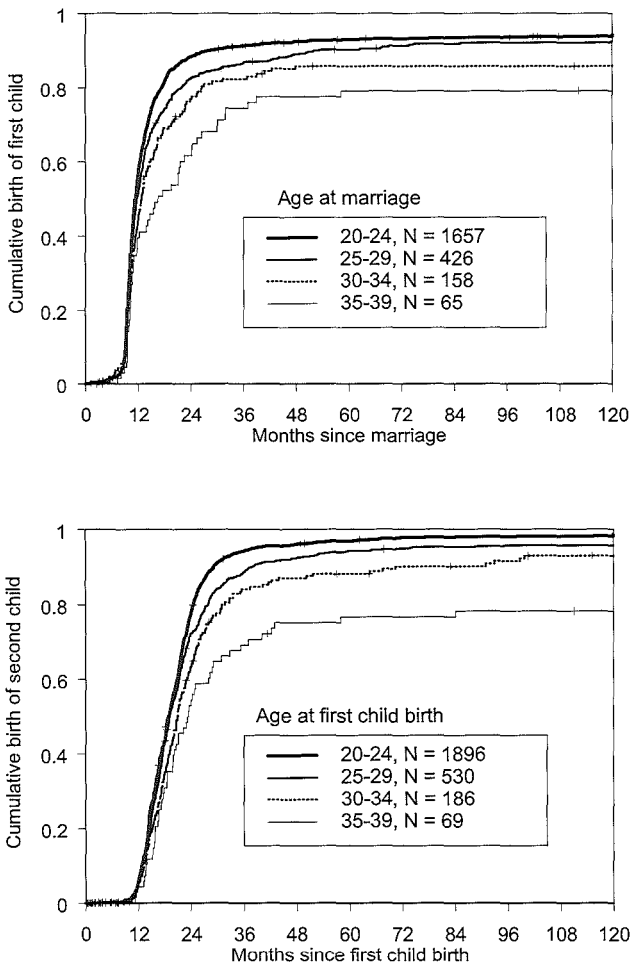


Figure 1 Cumulative probability of childbirth (Kaplan-Meier estimates for women from the Saguenay region, born between 1844 and 1899)

a) first childbirth, by age at marriage (n = 4,125)

b) second childbirth, by age at first childbirth (n = 3,761)

The percentage childlessness in this historical population (8.1% as a raw percentage, 7.4% estimated by the Kaplan-Meier method) is about twice as high as expected in a modern population (Greenhall and Vessey, 1990). Therefore, we adjusted the results for primary infertility by reducing childlessness by half, or, in formula form, by applying an extra fecundability ratio of 1.27 (= log(0.037) / log(0.074)) to the results of the Cox regression for primary infertility.

RESULTS

The Kaplan-Meier estimate of childlessness -corrected for under-representation of women with missing birth dates of children- was 7.4% after 10 years of marriage. This was 6.2%, 7.7%, 14.3% and 20.9% for women who were married at age 20-24, 25-29, 30-34 and 35-39 respectively (**Figure 1a**). The analysis was repeated for second childbirth (**Figure 1b**).

Ten years after the first childbirth, 3.2% of the primiparous women were estimated not to have experienced a second childbirth. This was 1.8%, 4.3%, 7.0% and 21.7% for women who had their first child at age 20-24, 25-29, 30-34 and 35-39 respectively. The relation between age at marriage and the percentage childlessness is depicted in more detail in **Figure 2**.

The results of the Cox regression analysis for the cohort, obtained by establishing a diagnosis of infertility are depicted in **Table 1**. A consistent downward trend in fecundability ratio is observed with increasing age and increasing duration of infertility.

Figure 2 Percentage childlessness by age at marriage for 4,125 women from the Saguenay region, born between 1844 and 1899. Estimates of a Cox regression model using a flexible spline function, with 95% Confidence limits.

Table 1 Fecundability ratio's of the Cox regression model for categories of age, duration of infertility and type of infertility. Estimated on data from a natural population.

	Fecundability ratio	95% Confidence interval
Age		
20-24 years [a]	1	-
25-29 years	0.83	0.70 – 1.00
30-34 years	0.67	0.51 – 0.87
35-39 years	0.45	0.31 – 0.64
Duration		
12 months [a]	1	-
24 months	0.49	0.45 – 0.54
36 months	0.37	0.33 – 0.42
48 months	0.26	0.22 – 0.31
Primary versus secondary infertility [a]	0.55 [b]	0.46 - 0.66

[a] Reference category

[b] The fecundability ratio of primary versus secondary infertility is adjusted upwards, corresponding with a reduction of the proportion childlessness in this population by half.

The 12 months prognosis calculated from this analysis is depicted in **Table 2,** for primary infertility (first column) and for secondary infertility (second column).

DISCUSSION

This study found a strong dependency of the chance of spontaneous pregnancy leading to live birth on female age, duration of infertility and parity, in a population that is likely to have experienced conditions of natural fertility, i.e. with negligible use of contraceptive practices.

Because clinical diagnosis of infertility is concerned with pregnancy and not with live birth, the cumulative Kaplan-Meier estimates for live birth were shifted back in time over a period of 9 months, giving the time of conception resulting in live birth. Still, this does not account for spontaneous abortions and miscarriages. Age-specific probabilities of foetal loss have been estimated to be 14%, 16%, 19% and 24% per recognised pregnancy at ages 25, 30, 35 and 40 respectively (Wood and Weinstein, 1988). In principle, these figures could be used to calculate back from first live birth to first pregnancy. However, this calculation is complicated because more than one pregnancy followed by foetal loss may have occurred before the pregnancy leading to live birth was realised. Back-calculation will result in higher cumulative pregnancy curves that are shifted back in time (Bongaarts, 1975). In the literature on fertility prognosis, an increasing awareness is emerging that the aim of infertility treatment is to achieve a live birth (Collins *et al.*, 1995 Snick *et al.*, 1997) and not just pregnancy. When clinicians adopt this point of view, we consider our results relevant for clinical practice.

Table 2 Cumulative probability of conception leading to childbirth (%) at 12 months after establishment of infertility, predicted by a Cox regression model with female age, duration and type (primary/secondary) of infertility.

Age at establishment of infertility [a]	Duration of infertility	Cumulative probability of conception leading to childbirth (% with 95% CI) at 12 months	
		Primary infertility	Secondary infertility [b]
20-24 years	12 months	44% (40-50)	63% (58-68)
	24 months	26% (22-28)	38% (34-43)
	36 months	20% (17-23)	31% (26-36)
	48 months	15% (12-17)	23% (19-28)
25-29 years	12 months	39% (33-44)	56% (51-62)
	24 months	21% (17-25)	33% (29-38)
	36 months	16% (14-20)	26% (23-31)
	48 months	12% (10-15)	20% (16-23)
30-34 years	12 months	32% (26-40)	48% (40-57)
	24 months	17% (14-22)	28% (22-34)
	36 months	14% (10-17)	22% (17-27)
	48 months	10% (8-12)	16% (12-20)
35-39 years	12 months	22% (16-32)	36% (27-47)
	24 months	12% (9-16)	19% (14-27)
	36 months	9% (6-12)	15% (11-21)
	48 months	6% (5-10)	11% (8-16)

[a] Means that the woman was aged, e.g., 20-24 years when her duration of infertility reached 12 months (or 24 months etc.)

[b] The start of the second child wish is assumed to be on average 6 months after the first childbirth. This was the average period of breastfeeding.

We defined secondary infertility as not having conceived a second child within a certain period following the first childbirth. The clinical definition of secondary infertility is different: it concerns also third, fourth and higher order conceptions, not just the second one. We note however that two-thirds of the births after the first child concerned second children in the year 2001 in the Netherlands (Centraal Bureau voor de Statistiek, 2002), so our definition covers a vast majority of secondary infertility in the modern clinical sense. Further, it seems plausible that the prognosis to conceive a following child is at least as good for women who already have more than one child as it is for women who have just one child. Therefore our results may be regarded as a lower limit of the prognosis of patients with secondary infertility.

The overall incidence of childlessness in this population was 7.4%, which is comparable to figures found in other historical natural populations (Leridon, 1977) (Menken *et al.*, 1986), but which is high in comparison with data on modern populations. A figure of 3-4% seems nowadays more appropriate (Greenhall and Vessey, 1990). Therefore we reduced the proportion childlessness by half in our

analysis of prognosis in primary infertility. Previously (te Velde *et al.*, 2000), we used a theoretical model for the heterogeneous distribution of monthly pregnancy chances in natural populations (Bongaarts, 1975), together with a 4% sterility rate. The resulting probability of conceiving within the next year was 49% after 1 year and 14% after 3 years. These figures are comparable to the results from the present study for women with primary infertility who are younger than 30 years (**Table 2**). Furthermore, the theoretical model predicts that 84% of women conceives their first child within one year after marriage (i.e. at duration zero), almost equal to the estimate of 86% for women aged 20-24 in our sample (**Figure 1a**). In the study by Snick on subfertile couples in a primary care setting (Snick *et al.*, 1997) patients were followed until spontaneous pregnancy or start of treatment. The 12 months prediction of pregnancy leading to live birth of patients with primary unexplained infertility ranged from 48% for women under 30 with duration of infertility ≤ 2 years to 27% for women over 30 and duration longer than 2 years. For secondary unexplained infertility the figures ranged from 63% to 42% respectively. The corresponding predictions for couples with a diagnosis of semen defect were 32% to 17% (primary) and 44% to 27% (secondary) respectively. The findings of the present study, in which no diagnosis is known, are in between these values.

The high proportion of childlessness in this historical population raises several questions about the validity of the extrapolation to modern times. Was this really a natural fertility population in the sense that no use was made of contraception, either to avoid getting any children (voluntary childlessness) or to limit family size? Was the incidence of infertility in the 19th century comparable to the present incidence? And finally, were women who married late in this on average very young marrying population comparable to women in modern times who voluntarily choose to start getting children late?

No data are available on voluntary childlessness, but the religious nature of the population (Roman Catholic) and the ongoing tradition of high fertility must have put social pressure on all couples to get children: the total fertility rate for the whole population of Quebec was 5.6 children per woman in 1891 (Henripin, 1968). One indication of the near absence of contraceptive use comes from an analysis of the age-specific average number of children per married woman per year. This was 0.5 child/year at ages between age 20 and 29 and steadily declined at older ages. If couples had wanted to limit their family size, women who were married at age 20 would have completed their family well before they were 40. They would have had a lower average number of children per year at 35-39 years than women who only married at age 35. However, the age specific average number of children per woman did not depend on the age of marriage of the woman (data not shown). Absence of methods to limit family size is further illustrated by the 'completed fertility' (i.e. the mean number of children at age 50 for women still married at this age) classified by age at marriage. It was 10.8 and 6.8 for women married before age 20 and at age 25-29 respectively. The high average number of children in the women who married at

young age is not compatible with a widespread use of methods to limit family size. Of course, this still doesn't exclude the possibility that a small and select group of couples remained voluntarily childless.

We have no data on the incidence of true sterility in this population, but we do have data on mortality that give some indications about the general health of men and women. The life expectancy at birth was 43 years for males and 41 years for females, while at age 20 the expected total duration of life was 65 and 60 years respectively. The low life expectancy at birth was due to childhood mortality: 32% of men and 30% of women died before the age of 10 years. The lower life expectancy of women compared to men was largely due to maternal mortality: overall 6.6% of women died during or within a month after the birth of a child. Indeed, the expected total duration of life for men and women who had reached age 50 was comparable: about 73 years for both. These figures suggest that the general health of this population wasn't that much poorer than that of a modern western society, but rather that infant mortality was high and obstetrical care was largely lacking. Infections associated with complications during childbearing are probably the main cause of maternal mortality. Even if a woman survived such an infection, her health status may have been compromised and her fertility hampered, e.g. through tubal occlusion caused by the infection. This may have influenced our results for secondary infertility, but it does not explain why so many women remained childless. Consanguineous marriages occurred, but not in a high proportion, as special dispensation was required from the Church. No differences in time to first or second childbirth were observed between consanguineous and non-consanguineous marriages (fecundability ratio consanguineous versus non-consanguineous: 1.09, 95% CI (0.91 to 1.31)).

The age at first childbirth in this historic population was considerably younger than it is at present in western societies. Nevertheless we extrapolated our findings to modern times. One may wonder whether a 19th century woman from Quebec Canada, who married at the late age of 30, really had the same pregnancy prospects as a woman in a modern western society such as the Netherlands, for whom 29.2 years is the average age on which the first child is born. However, we note that women who married at age 40, still had a 43% chance (95% CI 33% to 53%) of eventually getting a child (**Figure 2**). In this same population, the median age at last childbirth of women who married before the age of 25 was 41.4 years, i.e. 50% of the young-married women was able to get children after the age of 41. Thus the fertility of women who married at the late age of 40 was comparable to the fertility of young-married women at about that age.

Which lessons for modern clinical practice can be deduced from these historic data? In this population nothing is known about possible diagnoses of infertility. Therefore, the predicted prognosis for a certain age and duration of infertility applies only to the average of a group of patients with these characteristics. Our results indicate that after 1 year of primary infertility in patients under 35, and 1-2 years of secondary infertility irrespective of age, prospects are on average still good, since 30% or more of patients

will become pregnant within the next year. If diagnostics are started at these durations of infertility (earlier on indication) and nothing is found, we may conclude that prognosis is better than the population average, and therapy could be withheld for the next 1 or 2 years. For young women (age < 30) with secondary infertility prospects are still considerable even after 3 years of infertility.

References

Bongaarts J. (1975) A method for the estimation of fecundability. *Demography.* 12, 645-60.

Bouchard G, Roy R, Casgrain B, Hubert M. (1989) Fichier de population et structures de gestion de base de données: le fichier-réseau BALSAC et le système INGRES/INGRID. *Histoire & Mesure.* 4, 39-57.

Centraal Bureau voor de Statistiek. (2002) Geboorte 1997-2001. *Maandstatistiek Bevolking.* 12, 26-27.

Collins JA, Burrows EA, Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril.* 64, 22-8.

Eimers JM, te Velde ER, Gerritse R, Vogelzang ET, Looman CW, Habbema JD. (1994) The prediction of the chance to conceive in subfertile couples. *Fertil Steril.* 61, 44-52.

Greenhall E, Vessey M. (1990) The prevalence of subfertility: a review of the current confusion and a report of two new studies. *Fertil Steril.* 54, 978-83.

Henripin J. Tendances et facteurs de la fécondité au Canada. Ottawa: Bureau Fédéral de la Statistique, 1968:30.

Huber PJ. (1967) The behavior of maximum likelihood estimates under non-standard condition. *Proceedings Fifth Berkely Symposium Mathematical Statistics.* 1, 221-233.

Lalou R. La haute fécondité des Québécoises avant la révolution contraceptive: effet de nature ou effet de culture? Un bilan des connaissances. Chicoutimi: Université du Québec à Chicoutimi, 1990.

Leridon H. (1977) Sur l'estimation de la stérilité. *Population.* 32, 231-48.

Menken J, Trussell J, Larsen U. (1986) Age and infertility. *Science.* 233, 1389-94.

Snick HK, Snick TS, Evers JL, Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod.* 12, 1582-8.

te Velde ER, Eijkemans R, Habbema HD. (2000) Variation in couple fecundity and time to pregnancy, an essential concept in human reproduction. *Lancet.* 355, 1928-9.

Wood JW, Weinstein M. (1988) A model of age-specific fecundability. *Population studies.* 42, 85-113.

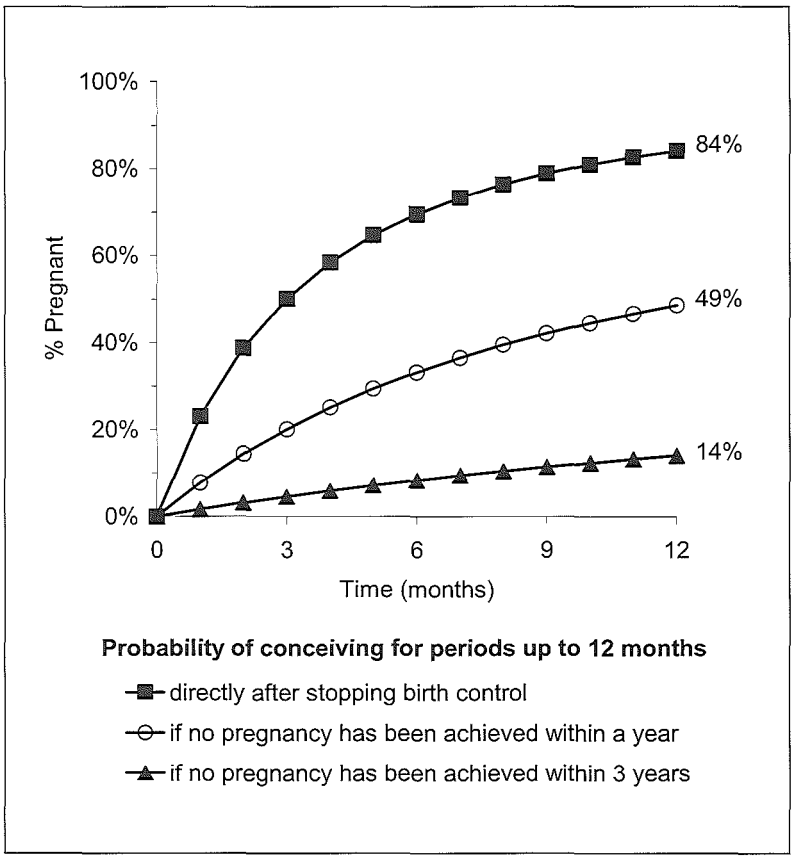
4

Variation in couple fecundity and time to pregnancy, an essential concept in human reproduction

In clinical practice infertility is usually defined as a failure to become pregnant during a 12-month period of regular, unprotected intercourse. Hence fertility is assumed to be normal when a pregnancy occurs within this period. This definition has led to a dichotomous concept -of fertility or infertility-, which not only ignores the essence of reproduction, but also gives rise to misinterpretations. For example, couples are inclined to think that, if a pregnancy has not occurred within a year, they are sterile and require immediate infertility treatment.

In fact, reproduction is a matter of chance depending on the subtle balance between success or failure of complex, mostly poorly understood, sequential processes that may lead to a pregnancy and eventually to the birth of a healthy child. These processes include spermatogenesis and oogenesis, sexual intercourse and transport of gametes, fertilisation, migration of the embryo to the uterus and its subsequent implantation, and finally intrauterine development of the fetus. Failure can occur at any link of this delicate chain, but most commonly does so at the early stages. With regular intercourse a new chance of pregnancy arises every menstrual cycle. Demographic studies show that the distribution of monthly fertility of couples trying to conceive is heterogeneous and fits a beta-distribution (Bongaarts, 1975), (Leridon and Spira, 1984). Each couple has a more or less constant monthly probability of conceiving, but between couples the probabilities vary widely, from 0% to an upper limit of about 60%. A monthly fertility of zero corresponds to true infertility in the sense of sterility, which occurs in 3-5% of all couples (Greenhall and Vessey, 1990).

With a high monthly fertility the average time to pregnancy is short, and vice versa (Baird *et al.*, 1986). Survival analysis is the appropriate method for assessing time-to-pregnancy data. Since the more-fertile couples tend to conceive first, as time goes by progressively less-fertile couples selectively remain in the population of couples who have not achieved a pregnancy (Leridon, 1977). Hence how long couples have been unsuccessful at conceiving is an essential estimate of the degree of subfertility. The figure, which is derived from data collected by Bongaarts in natural, non-



contraception-practising populations (Bongaarts, 1975) and based on a sterility rate of 4%, shows that, as the duration of inability to achieve a pregnancy increases, the probability of success during the subsequent year sharply decreases. However, the data also show that, if the unproductive period is short (eg, 12 months), the probability of success is still considerable -half the couples will conceive during the following year. These reasonable pregnancy prospects illustrate that the definition of infertility as a failure to conceive within a year contains an oversimplification that may result in premature resort to assisted-reproduction techniques, with their associated risks.

Time to pregnancy has also been used in reproductive epidemiology for assessing exposures related to lifestyle, environment, or occupation in case-control studies (Baird *et al.*, 1986), (de Cock *et al.*, 1994). If time to pregnancy is significantly longer in the exposed couples than in the controls, the conclusion that the exposure negatively affects fertility is justified. In several studies time to pregnancy seems to be a sensitive outcome measure for overall couple fertility. It does not enable conclusions as to which link of the reproduction chain is affected. Other disadvantages are that time to pregnancy in this type of studies is susceptible to bias and confounding, and that it

evaluates only exposures in couples who sooner or later achieve a pregnancy; it cannot assess the effect on the proportion of truly infertile couples (Baird *et al.*, 1986)

Determinants of sperm quality such as sperm concentration are other measures used in reproductive epidemiology (Bonde and Giwercman, 1995). Several studies have shown a secular trend of decline in these variables since the fifties (Carlsen *et al.*, 1992), (Auger *et al.*, 1995). Because this finding may be a reflection of the effect of environmental pollution, it has raised serious concern (Sharpe and Skakkebaek, 1993). In a paper in today's Lancet, Michael Joffe started with the hypothesis that a decline in male fertility, with an increase in time to pregnancy, is to be expected as a consequence of declining sperm quality (Joffe, 2000).

Instead, he reports a significant decrease in time to pregnancy over the past 40 years. He concludes that, if a decline in male fertility has occurred, it has been fully compensated by other factors related to couple fertility. Another conclusion should be that at present the near panic sometimes expressed in the lay press about the effects of environmental pollution on sperm quality and male fertility is not justified.

References

Auger J, Kunstmann JM, Czyglik F, Jouannet P. (1995) Decline in semen quality among fertile men in Paris during the past 20 years. *N Engl J Med.* 332, 281-5.

Baird DD, Wilcox AJ, Weinberg CR. (1986) Use of time to pregnancy to study environmental exposures. *Am J Epidemiol.* 124, 470-80.

Bonde JP, Giwercman A. (1995) Occupational hazards to male fecundity. *Reproductive Med Rev.* 4, 59-73.

Bongaarts J. (1975) A method for the estimation of fecundability. *Demography.* 12, 645-60.

Carlsen E, Giwercman A, Keiding N, Skakkebaek NE. (1992) Evidence for decreasing quality of semen during past 50 years. *Bmj.* 305, 609-13.

de Cock J, Westveer K, Heederik D, te Velde E, van Kooij R. (1994) Time to pregnancy and occupational exposure to pesticides in fruit growers in The Netherlands. *Occup Environ Med.* 51, 693-9.

Greenhall E, Vessey M. (1990) The prevalence of subfertility: a review of the current confusion and a report of two new studies. *Fertil Steril.* 54, 978-83.

Joffe M. (2000) Time trends in biological fertility in Britain. *Lancet.* 355, 1961-5.

Leridon H. (1977) *Human fertility: the basic component*, Chicago University Press,Chicago.

Leridon H, Spira A. (1984) Problems in measuring the effectiveness of infertility therapy. *Fertil Steril.* 41, 580-6.

Sharpe RM, Skakkebaek NE. (1993) Are oestrogens involved in falling sperm counts and disorders of the male reproductive tract? *Lancet.* 341, 1392-5.

Part II

Prediction of treatment outcome in anovulatory infertile patients

5

Introduction Part II: Prediction of ovulation induction outcome in normogonadotropic anovulatory infertility

BACKGROUND

In the Netherlands, each year about 13,000 couples visit a fertility clinic for the first time because of an unresolved child wish. In 30% of cases the diagnosis is "cycle disturbances" (i.e. about 4,000 women) (Rowe *et al.*, 1993). There are many causes of anovulation that result in cycle disturbances. The usual clinical classification, recommended by the World Health Organization (Rowe *et al.*, 1993), is based on serum levels of hormones. Low gonadotrophins together with low oestrogen levels are suggestive of a central cause of disease at the hypothalamic-pituitary level (category World Health Organisation I (WHO I)), while low oestrogen levels together with elevated gonadotrophins are indicative of a disease at the ovarian level (WHO III). The most prevalent form however (80%, which amounts to 3,200 new patients per year in the Netherlands) is characterised by normal gonadotrophins together with normal oestrogen levels (WHO II). For this form of cycle disturbances several treatment options are available. Modern treatment options include laser ovarian surgery (Cohen, 1996) (Farquhar *et al.*, 2001) insulin sensitizing agents (Heard *et al.*, 2002) (Nestler *et al.*, 2002) (Glueck *et al.*, 2002) and aromatase inhibitors (Mitwally and Casper, 2001). However, most patients receive the classical treatment algorithm, consisting of three main treatment options. The treatments have different burden, risks, costs and chances of ongoing pregnancy. There are also huge differences between individual patients in response to these treatments. The current sequence of treatment is from the least burdensome and cheapest to the most burdensome and most costly treatment.

Almost all patients with WHO II anovulatory infertility receive, according to the conventional treatment algortihm, the anti-estrogen Clomiphene Citrate (CC) as first line treatment. CC is relatively cheap and has few side effects. The standard schedule is to administer CC during 5 days, in doses increasing from 50 mg during the first

cycle, via 100 mg to a maximum of 150 mg, depending on the occurrence of ovulation. About 25% of patients show no ovulation, even after the maximal dose: they have Clomiphene Resistant Anovulation (CRA). Of the women who do ovulate with CC, about 40% are still not pregnant after 6 - 9 months (Clomiphene Failures). Therefore, CC fails in 55% (= 25% + 0,4 x 75%) of patients, in spite of having been treated during half a year (Gorlitsky *et al.*, 1978) (Hammond *et al.*, 1983).

The women that do not become pregnant with CC are subsequently treated with Gonadotrophins. This is a more burdensome and also more expensive treatment than CC. The percentage ovulation per cycle of Gonadotrophin Ovulation Induction (GOI) is about 70% and the cumulative pregnancy rate after 7 cycles of treatment is at least 55% (van Santbrink *et al.*, 1995) (White *et al.*, 1996). The amount of follicle stimulating hormone (FSH) administered should, considering the individual ovarian response, be adjusted per patient. In doing this, one easily gives too much or too little because the therapeutical bandwidth of this medication is tight. Sometimes a GOI-cycle results in hyper response of the ovaries, with a large number of follicles growing at the same time. When treatment is continued in such a cycle, there is a risk of a large multiple pregnancy and/or of the potentially life threatening ovarian hyper stimulation syndrome (OHSS) (Fauser and Van Heusden, 1997). For these reasons, treatment in a cycle with hyper response should be cancelled. Despite intensive control during treatment, complications can never be ruled out, and some patients seem –in retrospect- to be more prone than others to get complications.

The last step in the treatment sequence for ovulation disorders is In Vitro Fertilisation (IVF). This is the most burdensome and most costly treatment. The risk of serious complications is however less than with GOI. There is a chance of a multiple pregnancy, but this chance is strongly determined by the number of embryos that is transferred to the uterus. The success rate of treatment is dependent on the pregnancy rate per cycle and the number of cycles that is performed. Recent figures for the Netherlands are: 19% per cycle with an average of 2.1 cycles, giving in total 40% ongoing pregnancies (Kremer *et al.*, 2002).

It seems obvious to deploy the least burdensome and cheapest treatment in the first line, and only after this turn to the more burdensome and more expensive treatments. Indeed, this is the current practice (Nederlandse Vereniging voor Obstetrie en Gynaecologie (NVOG) guidelines "Anovulatie en kinderwens", June 1996, composed by Fauser et al.).

However, it cannot be excluded that some patients could benefit from different treatment regimens than the standard. To find out which patients this concerns and what the alternative treatment regimens are, we need to know the chances of success for individual patients on each of these treatments, based on the standardised initial examination. For example, if we can predict beforehand that a patient has a very small chance to become pregnant with CC (= high chance of being CRA or Clomiphene

Failure), it might be more efficient to start immediately with GOI. In women of more advanced age this means that precious time – with fast declining pregnancy chances – is won. In case CC is the first line treatment, but doesn't lead to pregnancy, the GOI treatment –particularly the starting doses – may be geared to the CC response. This might diminish the chance of complications. Next, the decision to proceed to IVF might be partly based on the response during the CC and/or GOI cycles. It may be that for some patient groups, both CC and GOI could be skipped, based on the findings in the initial examination of the patient. Finally, there might be patient groups (e.g. with very mild cycle disturbances), whose chance on a spontaneous pregnancy is so high that treatment is not yet indicated.

RESEARCH QUESTIONS

In our quest for individualising treatment regimens for patients with WHO II anovulatory infertilily, we will address the following research questions.

1. **Which patient characteristics are predictive of pregnancy chances following ovulation induction with Clomiphene Citrate (CC)? What is -in this respect- the relation between the intermediate endpoint 'ovulation' and the final endpoint pregnancy?**

The prediction of pregnancy chances with CC treatment is methodologically challenging: patients who do not reach the intermediate endpoint 'ovulation' will stop CC-treatment, because their chances of becoming pregnant with it are zero. These patients cannot be analysed on the same level as the patients who *do* ovulate. Therefore, special attention will be devoted to the role of this intermediate (time-dependent) endpoint in the prediction of pregnancy chances with CC.

2. **Which patient characteristics are predictive of pregnancy chances of classical ovulation induction with Clomiphene Citrate (CC) followed by Gonadotrophin ovulation induction (GOI)?**

Part of the comparison of different treatment strategies is to be able to predict the pregnancy chances of the combined strategy of Clomiphene Citrate (CC) followed by Gonadotrophin ovulation induction (GOI) in of case absent pregnancy. Therefore, an overall analysis of the time to pregnancy with the combined strategy will be performed.

After these two, we are ready to address the final question:

3. **What is the optimal treatment sequence for WHO II patients based on their individual characteristics and on the response in previous treatment steps?**

The previously developed prediction models are used to perform a cost-effectiveness analysis for subgroups defined by individual patient characteristics that were found to

be predictive of outcome. For each subgroup, the cost-effectiveness of the standard treatment sequence CC -> GOI -> IVF will be compared to alternative treatment strategies in which one or more of the treatment steps are skipped.

OUTLINE OF PART II

In *chapter 6*, the methodology that is used to answer research question 1 is described, and the reasons are explained why the analysis was split into three parts.
The results described in *chapter 7* concern the prediction of the intermediate (time-dependent) endpoint Clomiphene Resistant Anovulation (CRA, i.e the definitive assessment that ovulation cannot be induced by CC in this patient). It forms the first step in answering research question 1.
The second step in answering research question 1 consists of the prediction of pregnancy chances in patients who are not CRA, and is presented in *chapter 8*.
Chapter 9 describes the third and final step in answering research question 1: the integration of the prediction model for the chance to be CRA with the prediction model for pregnancy chances in patients who are not CRA.
In *chapter 10* the analysis is presented that will answer research question 2, prediction of outcome after the combined sequence of CC followed by GOI.
In *chapter 11*: a cost-effectiveness analysis is performed to optimize the treatment strategy for anovulatory infertility (WHO II), based on individual patient characteristics. This chapter presents the final analysis, answering research question 3.

References

Cohen J. (1996) Laparoscopic procedures for treatment of infertility related to polycystic ovarian syndrome. *Hum Reprod Update.* 2, 337-44.

Farquhar C, Vandekerckhove P, Lilford R. (2001) Laparoscopic "drilling" by diathermy or laser for ovulation induction in anovulatory polycystic ovary syndrome. *Cochrane Database Syst Rev.* 4.

Fauser BC, Van Heusden AM. (1997) Manipulation of human ovarian function: physiological concepts and clinical consequences. *Endocr Rev.* 18, 71-106.

Glueck CJ, Wang P, Goldenberg N, Sieve-Smith L. (2002) Pregnancy outcomes among women with polycystic ovary syndrome treated with metformin. *Hum Reprod.* 17, 2858-64.

Gorlitsky GA, Kase NG, Speroff L. (1978) Ovulation and pregnancy rates with clomiphene citrate. *Obstet Gynecol.* 51, 265-9.

Hammond MG, Halme JK, Talbert LM. (1983) Factors affecting the pregnancy rate in clomiphene citrate induction of ovulation. *Obstet Gynecol.* 62, 196-202.

Heard MJ, Pierce A, Carson SA, Buster JE. (2002) Pregnancies following use of metformin for ovulation induction in patients with polycystic ovary syndrome. *Fertil Steril.* 77, 669-73.

Kremer JA, Beekhuizen W, Bots RS, Braat DD, van Dop PA, Jansen CA, et al. (2002) Resultaten van in-vitro fertilisatie in Nederland, 1996-2000. *Ned Tijdschr Geneeskd.*, 146, 2358-63.

Mitwally MF, Casper RF. (2001) Use of an aromatase inhibitor for induction of ovulation in patients with an inadequate response to clomiphene citrate. *Fertil Steril.* 75, 305-9.

Nestler JE, Stovall D, Akhter N, Iuorno MJ, Jakubowicz DJ. (2002) Strategies for the use of insulin-sensitizing drugs to treat infertility in women with polycystic ovary syndrome. *Fertil Steril.* 77, 209-15.

Rowe PJ, Comhaire FH, Hargreave TB, Mellows HJ. (1993) *WHO manual for the standardized investigation and diagnosis of the infertile couple.*, Cambridge University Press,Cambridge, UK.

van Santbrink EJ, Hop WC, van Dessel TJ, de Jong FH, Fauser BC. (1995) Decremental follicle-stimulating hormone and dominant follicle development during the normal menstrual cycle. *Fertil Steril.* 64, 37-43.

White DM, Polson DW, Kiddy D, Sagle P, Watson H, Gilling-Smith C, et al. (1996) Induction of ovulation with low-dose gonadotropins in polycystic ovary syndrome: an analysis of 109 pregnancies in 225 women. *J Clin Endocrinol Metab.* 81, 3821-4.

6

Characteristics of the best prognostic evidence: An example on prediction of outcome after Clomiphene Citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility

ABSTRACT

The standard first-line treatment for normogonadotropic anovulatory infertile patients [referred to as World Health Organization group 2 (WHO 2)] is ovulation induction using clomiphene citrate (CC) in incremental doses. Twenty to 25% of women show clomiphene-resistant anovulation (CRA), that is, they remain anovulatory even after multiple attempts with increased doses of CC. About 50% of the ovulatory CC patients conceive within six CC-induced cycles. Given the heterogeneous nature of the group, the individual prognosis (i.e., the chance of success) will vary considerably between patients. In the event an individual prognosis of each patient would be available before the start of the treatment, the overall efficiency of ovulation induction could be improved.

Prognostic evidence at an individual level should use multiple patient variables, including results from previous treatments (if any). When variables are interdependent, a statistical model can be used to relate individual characteristics with the predicted outcome. Such a model will provide estimates of prognosis for individualized patient profiles, allowing new patients to profit from the experience of the cohort of previous patients used to build the model. This paper discusses the prediction of time to pregnancy following induction of ovulation with CC. This prediction was broken down in two steps, leading to two separate prognostic models. The first model predicts an intermediate outcome, the chance that the patient will be CRA (i.e., no ovulation in response to CC medication); the second model predicts the final outcome (time until pregnancy) in women who do ovulate.

The CRA model was based on a prospective cohort study of 201 patients with normogonadotropic oligoamenorrheic infertility, 45 of whom were CRA (22%). It contained four predictor variables all related to the diagnosis of PCOS within the

group of WHO 2: Increased free androgen index (FAI; hyperandrogenemia), elevated body mass index (BMI; obesity), greater mean ovarian volume (as an ultrasound feature of polycystic ovaries), and amenorrhea were all predictive for CRA. The second model was based on the non-CRA patients and contained two prognostic variables: increased age and oligomenorrhea were predictive for longer time to pregnancy after first ovulation with CC. Using the example of the prediction of time to pregnancy following induction of ovulation with CC, we present and discuss characteristics of good prognostic evidence for clinical use, focusing on study design, statistical analysis, evaluation, and presentation of results.

INTRODUCTION

Anovulation is a frequent cause of infertility. These women usually present with oligo- or amenorrhea. In the great majority of infertile women presenting with cycle abnormalities, serum follicle-stimulating hormone (FSH) and estradiol (E2) levels are within normal limits. The group of patients presenting with normal FSH and E2 are referred to as World Health Organization group 2 (WHO 2) indicating a "pituitary–ovarian disbalance." Women within this heterogeneous group may (in addition to anovulation) present with obesity and hirsutism. Various endocrine abnormalities such as elevated luteinizing hormone (LH) levels, hyperandrogenemia, and insulin resistance, along with abnormalities in the insulin like growth factor system, have been observed. Moreover, polycystic ovaries may be found. Among many others, our group has focused on pathophysiology (Pache *et al.*, 1991) (van Dessel *et al.*, 1996) (Laven *et al.*, 2001), endocrine abnormalities (Fauser *et al.*, 1991), and ultrasound features (Pache *et al.*, 1992) involved in the polycystic ovary syndrome (PCOS). Depending on inclusion criteria used, between 40 and 70% of WHO 2 patients can be diagnosed as PCOS (van Santbrink *et al.*, 1995). This poorly defined syndrome is believed to be quite frequent among the female population (incidence around 5%) but little is known regarding underlying pathophysiological mechanisms.

The standard treatment for WHO 2 patients is ovulation induction, including clomiphene citrate (CC) in incremental doses as first-line therapy, followed by exogenous FSH in case of failure to ovulate or conceive. Twenty to 25% of women show clomiphene-resistant anovulation (CRA), that is, they remain anovulatory even after multiple attempts with increased doses of CC. About 50% of the ovulatory CC patients conceives within six CC-induced cycles. Of the women who fail to conceive with CC treatment, approximately 90% will ovulate and 50% will conceive subsequently following FSH ovulation induction. However, ovarian hyperstimulation and multiple pregnancies are frequent complications (Fauser and Van Heusden, 1997).

The above-mentioned figures apply to the group of WHO 2 patients as a whole. For an individual patient the outcome will be either success or failure. Given the heterogeneous nature of the group, the individual prognosis (i.e., the chance of success) will vary considerably between patients. If an individual prognosis of each patient were available before the start of the treatment, the overall efficiency of ovulation induction might be improved. A more patient-tailored treatment strategy could be designed, such as altered dose regimens, skipping first-line treatment with CC, or going directly in selected patients to alternative treatment modalities such as metformin or laser surgery of ovaries.

For subgroups defined by a single characteristic (e.g., amenorrhea or age > 30), prognostic evidence may come from experience in large groups of patients with this characteristic. Prognostic evidence at an individual level, however, should use multiple patient variables, including results from previous treatments (if any). When variables are interdependent, a statistical model can be used to relate individual characteristics

with a predicted outcome. Such a model will provide estimates of prognosis for individualized patient profiles, allowing new patients to profit from the experience of the cohort of previous patients used to build the model.

Throughout this paper we will discuss the prediction of time to pregnancy following induction of ovulation with CC. This prediction was broken down into two steps, leading to two separate prognostic models. The first model predicts an intermediate outcome, the chance that the patient will be CRA (i.e., no ovulation in response to CC medication) (Imani *et al.*, 1998); the second model predicts the final outcome (time until pregnancy) in women who do ovulate (Imani *et al.*, 1999). The CRA model was based on a prospective cohort study of 201 patients with normogonadotropic oligoamenorrheic infertility, 45 of whom were CRA (22%). It contained four predictor variables all related to the diagnosis of PCOS within the group of WHO 2: increased free androgen index (FAI; hyperandrogenemia), elevated body mass index (BMI; obesity), greater mean ovarian volume (as an ultrasound feature of polycystic ovaries), and amenorrhea were all predictive for CRA. The second model was based on the non-CRA patients and contained two prognostic variables: increased age and oligomenorrhea were predictive for longer time to pregnancy after first ovulation with CC.

Although the prognostic evidence contained in a statistical model will provide a better quantification of the relationship between predictors and outcome compared with informal clinical reasoning, it may perform disappointingly in clinical practice because of possible unreliability of the model predictions. Therefore, a study aimed at obtaining prognostic evidence should also assess the reliability of the evidence, through proper evaluation of the statistical model.

We will present and discuss characteristics of good prognostic evidence for clinical use, focusing on study design, statistical analysis, and evaluation and presentation of results.

STUDY DESIGN

The essence of prognosis is to relate patient and disease characteristics known at present to outcome in the future. Therefore, longitudinal data from a cohort study are required to construct such a prognostic model. The cohort study may be either prospective or retrospective. From a methodological point of view the prospective design is to be preferred, as it allows investigators to standardize treatment, guarantees that variables of interest are registered, minimizes missing data, and maximizes completeness of follow-up. In the design phase of the study decisions have to be made on the outcome measure(s) and on the potential predictive variables, based on prior information from literature and pathophysiological knowledge.

The CC study was a prospective longitudinal follow- up study, initiated in 1993. Consecutive patients who were referred by their general practitioner and diagnosed as

WHO 2 were included. The study was set up with several endpoints in mind, because it was aimed at longitudinally following patients throughout the various treatment steps (Imani *et al.*, 1998) (Imani *et al.*, 1999) (Imani *et al.*, 2000) (Imani *et al.*, 2002b) (Imani *et al.*, 2002a). The final outcome measure was time to first ongoing pregnancy. An important intermediate outcome in CC treatment was CRA. For the CC study, follow-up ended with the definitive assessment of CRA, with first pregnancy following CC, or with no pregnancy following at least one CC-induced ovulatory cycle. Initial screening characteristics together with anamnestic variables, such as age of the woman, duration of infertility, and prior pregnancies, were recorded in a database because they were potentially predictive of outcome. The screening characteristics were chosen with the presumed etiology of PCOS in mind, including cycle history, body weight, serum LH, androgen and insulin levels, and ultrasound features of polycystic ovaries.

ANALYSIS

Data Considerations

The outcome that is to be predicted can be continuous (e.g., FSH response dose in gonadotropin induction of ovulation (Imani *et al.*, 2002a)), dichotomous (e.g., CRA [yes/no] following clomiphene citrate medication (Imani *et al.*, 1998) (Imani *et al.*, 2000), pregnancy [yes/no] in an in vitro fertilization [IVF] cycle (Templeton *et al.*, 1996)), or "censored time to event" (e.g., time to pregnancy after first ovulation with CC (Imani *et al.*, 1999), time to spontaneous pregnancy or live birth (Eimers *et al.*, 1994) (Collins *et al.*, 1995) (Snick *et al.*, 1997), and cumulative pregnancy rates in IVF (Stolwijk *et al.*, 1996)). An observation on time to pregnancy is called "censored" if follow-up ended before pregnancy occurred. Preferably, coding of predictor variables is as detailed as possible (i.e., continuous if the underlying phenomenon is measured on a continuous scale). Categorizing of variables leads to loss of information and should be performed, if at all, in the analysis phase only.

Choice of Model

Several types of statistical techniques may be applied to relate predictor variables to an outcome. Among others we may mention regression models, neural networks (Cross *et al.*, 1995) (Baxt, 1995), classification and regression trees (Breiman *et al.*, 1984), and probabilistic networks (Pearl, 1988) (Heckerman *et al.*, 1992) (Shwe *et al.*, 1991). We will focus on regression models, because they are most frequently used in medical applications.

A regression model relates an outcome variable (Y) to one (univariable) or to the weighted sum of several (multivariable) predictor variables ($X = \{X_1 \ldots X_p\}$). The weights by which the variables are multiplied are called "regression coefficients" (β_i, i

= *1...p*) and represent the strength of the association between a predictor and the outcome. The weighted sum is called the prognostic index (PI). In formula:

$$PI = \beta_1 X_1 + \ldots + \beta_p X_p$$

Three main types of regression models should be distinguished, dependent on the type of outcome data (**Table 1**): (1) ordinary linear regression for continuous outcomes, (2) logistic regression for dichotomous (or binary) outcomes, and (3) Cox proportional hazards regression for censored time to event data (also known as survival data).

The CRA prediction model was of type (2) logistic. The outcome CRA (yes or no) was known for all patients, because they were followed until first ovulation or definitive assessment of CRA. In contrast, the outcome for the time to pregnancy in women who ovulated following CC was not known for all patients: many patients had censored observations because their follow-up ended without pregnancy having occurred. But they could have become pregnant after their end of follow-up. The appropriate prediction model in women who ovulated following CC was of type (3), Cox proportional hazards regression.

In a univariable model, the magnitude of a regression coefficient may or may not be explained by the presence of one or more predicting variables that are not in the

Table 1: Three types of outcome data, with corresponding regression technique and interpretation of regression coefficients

Type of outcome Y Regression technique	Relationship between outcome (*Y*) and prognostic index (*PI*)	Interpretation of regression coefficient β_i
Continuous linear	$Y\|X = \beta_0 + PI$ [a,b]	β_i: change in average outcome per unit change of the variable
Dichotomous Logistic	$\Pr\{Y = 1 \mid X\} = \frac{1}{1 + Exp[-(\beta_0 + PI)]}$ [a,b,c]	$Exp(\beta_i)$ = Odds Ratio (OR): change in Odds on outcome per unit change of the variable
Censored time to event Cox proportional hazards	$\Pr\{\text{free of event at } Y = T \mid X\} = S(Y = T \mid X) = S(Y = T \mid 0)^{Exp(PI)}$ [b,d]	$Exp(\beta_i)$ = Hazard ratio (HR): change in hazard on outcome per unit change of the variable

[a] β_0 is the intercept of the regression formula

[b] $PI = \beta_1 X_1 + \ldots + \beta_p X_p$

[c] $\Pr\{Y=1|X\}$ is the probability that the dichotomous outcome Y will take the value 1, for a patient with predictor variables X.

[d] $\Pr\{\text{free of event at } Y=T|X\} = S(Y = T|X)$ is the probability that the event has not yet occurred at follow-up time $Y = T$, for a patient with predictor variables X (this is also known as the Survival probability, referring to the origins of this method that lie in analysis of mortality). $S(Y = T|0)$ is the 'baseline' survival probability, corresponding to a patient with all predictor variables equal to 0. Note that for each follow-up time T the survival probability can be calculated, giving a survival curve. When the event of interest is pregnancy instead of mortality the 'one minus survival' curve is usually preferred.

model. In multivariable models, regression coefficients are corrected for the prognostic contribution of the other variables in the model. In the analysis for CRA, a total of 9 of 16 potential predictor variables were significant in univariable analysis, but only four were left in the multivariable model. The prognostic contribution of the other five univariably significant variables became negligible and statistically nonsignificant after correction for the four variables in the final model. For instance, hyperandrogenemia was univariably significant but made no prognostic contribution to the multivariable model with FAI in it.

Informative Censoring

Cox regression makes the assumption that censoring is uninformative. In our time to pregnancy analysis this means that the fact that the follow-up of the patient ended before the patient had become pregnant (censoring) is not related to her pregnancy chances. Our initial objective was to construct a single prognostic model for pregnancy following CC-medication, using Cox regression. However, we realized that patients who dropped out because they became CRA violated the assumption of uninformative censoring. Because these patients didn't ovulate, they had a very low, if not zero, probability of conceiving. Assuming that these patients had the same probability of conceiving as patients who did not drop out is wrong. For this reason Cox regression could not be applied directly. This was the reason for the approach taken, in which separate analyses were performed for the chance to reach ovulation (i.e., predicting who will become CRA (Imani *et al.*, 1998)) and for the time until pregnancy, given that ovulation had occurred (Imani *et al.*, 1999). Logistic regression was used to predict who became CRA. For the analysis of time to pregnancy in case of ovulation, Cox regression was used in the non-CRA patients, with possible censoring at the end of their follow-up. Reasons for censoring could be dropout because the patient moved to another residence or because the couple had marital problems. We assumed that dropout for these reasons was not related to pregnancy chances and therefore this censoring was uninformative. Some patients were censored because they started CC treatment just a few months before the end date of the study (December 1997). Patients who were not pregnant at this time were also censored. In general, censoring because of study design or protocol is uninformative, because no prognosis-related selection of patients occurs.

Model Development

Determining which predictor variables will be part of the model is a major challenge. Already in the design phase of the study choices have been made on which variables to register. Usually a wide scope of variables is included so as to not miss factors that could later appear to be important. Especially in anovulation (a medical field in which the pathophysiology of disease is not completely resolved and many new factors are reported continuously) it is tempting to include many variables. For prognostic modeling this is usually not sensible. Including all registered variables as predictors in

the prognostic model may lead to serious overfitting and also to redundancy in case of mutually strongly correlated predictors. Furthermore, it is impractical to use a model with many predictors. Often, a reduction in the number of predictor variables is required.

Preferably, a conservative approach should be used in model reduction: start with a limited set of variables that have been established as having prognostic value in the literature. It is generally advised to use no more than one potential predictor on every 10 cases (i.e., patients in ordinary linear regression or events in logistic and Cox regression) (Harrell *et al.*, 1985). Experience in large simulation studies shows that a model with this approach will perform well in a new, independent data set (Steyerberg *et al.*, 2000). Often, further reduction of the number of predictors is achieved by applying statistical criteria such as backward or forward stepwise selection. These methods delete or include variables in a stepwise fashion according to repeated statistical significance testing. In this way noninformative variables are excluded from the model, but the danger is that some informative variables are also excluded. Furthermore, it involves statistical testing and fitting on the same data, known to lead to biased estimates of the coefficients for the variables that are included (Miller, 1990) (Hurvitch and Tsai, 1990). Simulation studies have shown that stepwise selection with the usual significance level of $P = 0.05$ may lead to poor model performance in new data. It is better to apply a less strict P value: P values of 0.10, 0.20, and even 0.50 have been proposed (Steyerberg *et al.*, 1999). In our case study $P = 0.10$ was used. For the CRA model we started with six plausible predictor variables that were significant in univariable analysis (amenorrhea, bleeding interval in case of oligomenorrhea, BMI, hyperandrogenemia, FAI, and mean ovarian volume) plus two variables that have appeared in literature as potential predictors (LH levels and mean follicle number). Using a stepwise backward elimination method four variables were retained in the final model: FAI, BMI, amenorrhea, and mean ovarian volume (**Table 2**).

Missing Data

Missing data may form a serious problem in multivariable modeling for several reasons. When the outcome variable is missing, bias may occur if values are missing selectively. For instance, women who become pregnant may be more willing to respond to a questionnaire on the outcome of treatment than women who don't get pregnant. As a result, the chance of pregnancy will be overestimated when only the women who responded are used for analysis. Care should be taken to avoid this as much as possible. When a predictor variable is missing, usually no bias is introduced when the couple is dropped from the analysis. However, it is a waste of data to drop a couple in which the outcome and many predictor variables have been measured, just because one of the predictor variables is missing. Imputation techniques may be used to keep these cases in the analysis (Little, 1992). Although imputation doesn't produce new independent information, it prevents existing information from being dropped.

Table 2: Univariable and multivariable (stepwise selected) Odds ratios in logistic regression predicting CRA.

Predictor variable	Range	Odds ratios (OR)[a] for CRA		Multivariable with Shrinkage correction [c]
		Univariable	Multivariable	
Clinical				
Amenorrhea vs oligomenorrhea	0 - 1	3.7**	5.0**	3.8
Bleeding interval in case of oligomenorrhea (days)	35 - 150	1.02*		
BMI (kg/m^2)	14 - 44	1.12**	1.10**	1.08
Endocrine				
Hyperandrogenemia (elevated T and/or AD)	0 - 1	3.24**		
FAI [b]	.5 - 24.2	1.25**	1.16**	1.13
LH (IU/L)	1.2 - 22.6	1.02		
Ultrasound				
Mean ovarian volume (mL)	2.7 - 22.1	1.16**	1.11*	1.09
Mean follicle number	3.0 - 25.0	1.06		

* $P < 0.05$

** $P < 0.01$

[a] The interpretation of Odds ratios in univariable analysis is as follows: Amenorrheic women are almost 4 times as likely (OR = 3.7) to be CRA as oligomenorrheic women. With each 1 point increase in BMI, the likelihood of being CRA increases (relatively) by 12% (OR = 1.12).

[b] Free androgen index = T × 100 / SHBG

[c] The multivariable odds ratios with shrinkage correction are calculated as $\text{Exp}(s \times \beta_i)$, with s = 0.82, the shrinkage factor and β_i the regression coefficient for variable X_i from logistic regression.

In very special circumstances the predictor variable is missing for a reason that is connected to selective values of the outcome variable. An example in our case study is formed by sperm parameters in predicting time to pregnancy. Some men didn't show up at the appointment for semen sampling because their wife became pregnant soon after start of medication. Probably the semen was of good quality in these cases, and the fact that the semen variables were missing was related both to the outcome variable (short time to pregnancy) and to the value of the semen variables. This is called "informative missingness" for which a correction should be applied. This has been done in our analysis (Imani *et al.*, 1999). Using the imputation for missing semen data, the Cox regression model for time to pregnancy in patients who were non-CRA was build in a stepwise fashion, starting with the seven variables that had $P \_ 0.20$ in univariable analysis (age of the woman, amenorrhea, bleeding interval in case of oligomenorrhea, LH, FAI, androstenedione, and estradiol). In the final model only two variables (age and amenorrhea) were retained. Both forward and backward variable selection produced the same final model.

EVALUATION: GENERALIZABILITY/ VALIDATION

The purpose of developing a prognostic model is to provide valid predictions in future patients, and it is the task of the developers of the model to assess its validity. Validity refers to a number of concepts that all relate to whether the model predictions can be trusted:

- Random or prediction error: How precise are the predictions?
- Systematic error or reliability: Do the predictions agree with observations? Has to do with model misspecification, or omitted predictor variables.
- Discrimination: To what extent are the model predictions able to separate between outcome categories?

When validity is assessed on exactly the same patients that were used to develop the model, we speak of "apparent validity." Validity in new but similar patients from the same setting is called "internal validity," and "external validity" refers to patients from another time or place. For assessment of external validity a new study should be set up. Preferably, this should be a prospective longitudinal cohort study, just like the study that was used to develop the model.

Usually, we can only assess apparent and internal validity. For apparent validity we simply apply the model to the same data that were used to develop the model. For internal validity we use the same data, but model development and assessment of validity are performed on different parts of the data. Cross-validation is a method in which the data set is randomly split in a number of parts of equal size. In turn, each part is used to evaluate a model that has been developed on the complementary parts. Well-known examples are the split-quarter and split-half methods. In the most extreme variant a single patient is left out to evaluate a model build on all other patients. This is then repeated for each patient (leave-one-out method).

The major drawback of cross-validation is that it is inefficient because only part of the data is used for model development or evaluation. A method that uses all available data is the bootstrap method. In this method the model-building process (selection of variables in the model and parameter estimation) is repeated a pre-specified number of times (e.g., 200 times). Each repetition consists of creating a new data set (bootstrap sample) by drawing cases with replacement from the original data. The resulting model from each bootstrap sample is evaluated on the original data. This procedure mimics what would happen if new data were collected repeatedly in the same setting (Efron and Tibshirani, 1993) and therefore assesses internal validity.

Random or Prediction Error

Uncertainty about the predictions of the model will always exist because of the statistical imprecision of the regression coefficients. For a given patient profile, the model produces a predicted outcome, with an associated 95% confidence interval. The prediction is to be interpreted as a mean value for this patient profile, and the

confidence interval refers to the uncertainty about this mean value. For dichotomous outcomes, such as in logistic regression predicting CRA, the prediction is to be interpreted as the proportion of patients with this profile that is expected (by the model) to be CRA.

Systematic Error or Reliability

Predictions of the model may also show systematic error when compared with actual observations. Reliability (or calibration) refers to how well predictions agree with observations. For example, if the model predicts that a patient has a 70% probability of being CRA, is the probability really 70%? Of course the real probability cannot be observed in an individual patient, because she either is CRA or not, but one could imagine a group of identical patients and determine their frequency of CRA. Assessing apparent reliability is equivalent to verifying that the model fits well to the data, by performing a goodness of fit test such as the Hosmer-Lemeshow for logistic regression (Hosmer and Lemeshow, 1989). Lack of fit may occur with improper modeling of the "dose-response" relationship between a continuous predictor and the outcome (e.g., the usual linear form is chosen where the data follow a nonlinear "bathtub" kind of curve), or because important interactions between predictors are omitted from the model. To resolve lack of fit, subject knowledge from pathophysiology may be of greater help than statistical "data dredging," because of the risk of chance findings and overfitting. This is particularly so when you are already asking much of the data (by considering too many variables, stepwise selection).

When there is no (more) lack of fit, the agreement between predictions and observations will be good in the data that were used to develop the model. However, statistical theory (Copas, 1983) (Van Houwelingen and Le Cessie, 1990) and recent simulation studies (Steyerberg *et al.*, 2000) have shown that in new patients the model predictions will be overoptimistic: higher than average predictions are too high and lower than average predictions are too low, a phenomenon related to "regression to the mean." It becomes worse when the number of candidate predictor variables is higher relative to the number of patients with the outcome. It has been shown that reliability in new patients will improve if all regression coefficients are corrected by a shrinkage factor (Van Houwelingen and Le Cessie, 1990). The bootstrap method is well suited to estimate the shrinkage factor that is required. Using the bootstrap method with 200 replications, we found a shrinkage factor of 0.82 for the CRA prediction model. This means that the final multivariable model, on average, will give the best fitting predictions in new data when all regression coefficients are 18% smaller in absolute size.

Discriminative Ability

Discrimination refers to the ability of the model to discriminate between good and poor prognosis patients, and it is a quantification of the degree to which predicted probabilities are lower for patients with poorer outcome. For logistic models it may be

measured by the area under the ROC curve (AUC), for Cox models by the c-statistic (Harrell *et al.*, 1984). They have a similar interpretation: given two randomly selected patients with different outcomes (e.g., one CRA, the other non-CRA or the time to conception of one patient is shorter than for the other one), AUC or c-statistic gives the probability that the model prediction is worse in the patient with the worst outcome. For sensible models it should be higher than 0.5, which is the AUC of a noninformative "flip of a coin" model.

Apparent discriminative ability is always optimistic. Internal validation techniques such as the bootstrap method will result in lower and more appropriate values for AUC or c-statistic than the apparent value. For the final CRA prediction model, the apparent AUC was 0.84, and the "optimism corrected" value, as determined by bootstrapping, was 0.82. Discrimination is an intrinsic quality of the model: two different models applied to one data set may have different discrimination. But it also depends on the range of outcome and predictor variables in the population.

In **Figure 1**, the distribution of predicted probability to be CRA is plotted for the ovulatory and CRA patients separately. The figure shows the amount of spread in prognosis the model has accomplished. The less overlap the two groups show, the easier it will be to distinguish ovulatory patients from CRA patients on the basis of the predicted probability, and the better the discrimination will be.

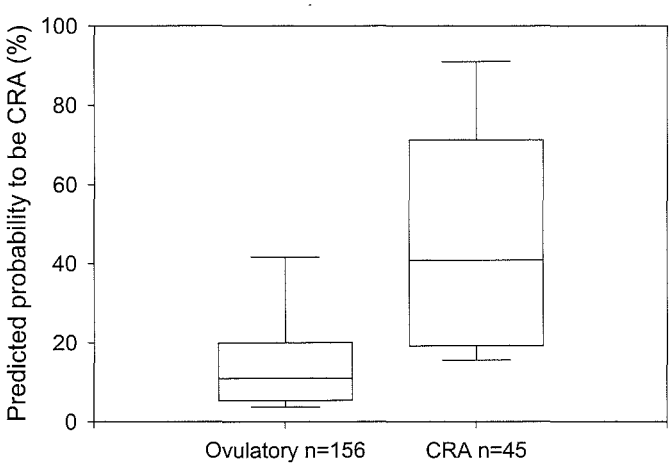


Figure 1: Box and whisker plot of the distribution of predicted probability to be CRA in ovulatory and in CRA patients respectively. Boxes represent median and 25th and 75th percentiles. Whiskers represent the 10th and 90th percentiles.

PRESENTATION

Finally, the results of model development and (internal) validation have to be presented. This may be done in the form of the regression formula, with the estimated coefficients, corrected by a shrinkage factor, obtained from the statistical software. Although the calculation of the prognostic index may be done with pen and paper, for logistic regression and the Cox model the transformation from prognostic index to predicted outcome variable (the "link function") is too complicated to perform by hand (Table 1), and users will have to implement the formula in computer software such as a spreadsheet. Alternatively, a score chart may be constructed. This consists of a score table, in which, for each value of the predictors, a score is assigned corresponding to the "shrunken" regression coefficient from the model. For this purpose, continuous predictors are divided into categories. The clinician has to look up the scores corresponding to the values of the predictor variables and add them, giving a sum score. Finally, the probability corresponding to the sum score can be read from a graph or table that also has to be provided.

Alternatively, a fully graphical presentation form, such as a nomogram, may be chosen. For the CRA prediction model, a special score chart was devised, containing the four predictors in the model (Imani *et al.*, 1998). To obtain an over all prediction of pregnancy chances, prior to start of CC medication (and thus prior to knowledge about CRA), the models for CRA and for prediction of pregnancy in case of ovulation were combined in a nomogram (Imani *et al.*, 2002b).

RECOMMENDATIONS

1. Choose the outcome parameter(s) carefully and pay due attention to the design of the study. In fertility medicine, a prospective cohort study is in almost all cases the most appropriate design.
2. Decide carefully which variables to include as potential predictors. This means in terms of anovulation which biologically plausible factors may be involved in the etiology of ovarian dysfunction and may predict outcomes of ovarian stimulation. Find the right balance between completeness (you don't want to miss important predictors) and parsimony (too many predictors will result in complex and badly performing models).
3. Choose the proper type of statistical analysis. Beware of violation of model assumptions, particularly when censored time to event such as time to pregnancy is the outcome parameter.
4. Perform internal validation by cross-validation or bootstrapping. This will result in a shrinkage factor for the prognostic index, as well as better assessment of model performance measures such as the AUC.

References

Baxt WG. (1995) Application of artificial neural networks to clinical medicine. *Lancet.* 346, 1135-8.

Breiman L, Friedmand L, Olshen R, Stone CJ. (1984) *Classification and regression trees*, Wadsworth,Belmont, CA.

Collins JA, Burrows EA, Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril.* 64, 22-8.

Copas JB. (1983) Regression, prediction and shrinkage (with discussion). *Journal of the Royal Statistical Society Series B.* 45, 311-354.

Cross SS, Harrison RF, Kennedy RL. (1995) Introduction to neural networks. *Lancet.* 346, 1075-9.

Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman and Hall, 1993.

Eimers JM, te Velde ER, Gerritse R, Vogelzang ET, Looman CW, Habbema JD. (1994) The prediction of the chance to conceive in subfertile couples. *Fertil Steril.* 61, 44-52.

Fauser BC, Pache TD, Lamberts SW, Hop WC, de Jong FH, Dahl KD. (1991) Serum bioactive and immunoreactive luteinizing hormone and follicle- stimulating hormone levels in women with cycle abnormalities, with or without polycystic ovarian disease. *J Clin Endocrinol Metab.* 73, 811-7.

Fauser BC, Van Heusden AM. (1997) Manipulation of human ovarian function: physiological concepts and clinical consequences. *Endocr Rev.* 18, 71-106.

Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. (1984) Regression modelling strategies for improved prognostic prediction. *Stat Med.* 3, 143-52.

Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA. (1985) Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep.* 69, 1071-77.

Heckerman DE, Horvitz EJ, Nathwani BN. (1992) Toward normative expert systems: Part I. The Pathfinder project. *Methods Inf Med.* 31, 90-105.

Hosmer DW, Lemeshow S. Applied logistic regression. New York, NY: John Wiley & Sons Inc, 1989:140-145.

Hurvitch CM, Tsai CL. (1990) The impact of model selection on inference in linear regression. *Am Statist.* 44, 214-217.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (1998) Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 83, 2361-5.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (1999) Predictors of chances to conceive in ovulatory patients during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 84, 1617-22.

Imani B, Eijkemans MJ, de Jong FH, Payne NN, Bouchard P, Giudice LC, et al. (2000) Free androgen index and leptin are the most prominent endocrine predictors of ovarian response during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 85, 676-82.

Imani B, Eijkemans MJ, Faessen GH, Bouchard P, Giudice LC, Fauser BC. (2002a) Prediction of the individual follicle-stimulating hormone threshold for gonadotropin induction of ovulation in normogonadotropic anovulatory infertility: an approach to increase safety and efficiency. *Fertil Steril.* 77, 83-90.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (2002b) A nomogram to predict the probability of live birth after clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *Fertil Steril.* 77, 91-7.

Laven JS, Imani B, Eijkemans MJ, de Jong FH, Fauser BC. (2001) Absent biologically relevant associations between serum inhibin B concentrations and characteristics of polycystic ovary syndrome in normogonadotrophic anovulatory infertility. *Hum Reprod.* 16, 1359-64.

Little RJA. (1992) Regression with missing X's: a review. *J Am Stat Assoc.* 87, 1227-1237.

Miller AJ. (1990) *Subset selection in regression*, Chapman & Hall,London.

Pache TD, Chadha S, Gooren LJ, Hop WC, Jaarsma KW, Dommerholt HB, et al. (1991) Ovarian morphology in long-term androgen-treated female to male transsexuals. A human model for the study of polycystic ovarian syndrome? *Histopathology.* 19, 445-52.

Pache TD, Wladimiroff JW, Hop WC, Fauser BC. (1992) How to discriminate between normal and polycystic ovaries: transvaginal US study. *Radiology.* 183, 421-3.

Pearl J. (1988) *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann Publishers,Palo Alto CA.

Shwe MA, Middleton B, Heckerman DE, Henrion M, Horvitz EJ, Lehmann HP, et al. (1991) Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. I. The probabilistic model and inference algorithms. *Methods Inf Med.* 30, 241-55.

Snick HK, Snick TS, Evers JL, Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod.* 12, 1582-8.

Steyerberg EW, Eijkemans MJ, Habbema JD. (1999) Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *J Clin Epidemiol.* 52, 935-42.

Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. (2000) Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med.* 19, 1059-1079.

Stolwijk AM, Hamilton CJ, Hollanders JM, Bastiaans LA, Zielhuis GA. (1996) A more realistic approach to the cumulative pregnancy rate after in- vitro fertilization. *Hum Reprod.* 11, 660-3.

Templeton A, Morris JK, Parslow W. (1996) Factors that affect outcome of in-vitro fertilisation treatment. *Lancet.* 348, 1402-6.

van Dessel HJ, Schipper I, Pache TD, van Geldorp H, de Jong FH, Fauser BC. (1996) Normal human follicle development: an evaluation of correlations with oestradiol, androstenedione and progesterone levels in individual follicles. *Clin Endocrinol (Oxf).* 44, 191-8.

van Houwelingen JC, Le Cessie S. (1990) Predictive value of statistical models. *Stat Med.* 9, 1303-25.

van Santbrink EJ, Hop WC, van Dessel TJ, de Jong FH, Fauser BC. (1995) Decremental follicle-stimulating hormone and dominant follicle development during the normal menstrual cycle. *Fertil Steril.* 64, 37-43.

7

Predictors of Patients Remaining Anovulatory during Clomiphene Citrate Induction of Ovulation in Normogonadotropic Oligoamenorrheic Infertility

ABSTRACT

The diagnostic criteria used to identify patients suffering from polycystic ovary syndrome remain controversial. The present prospective longitudinal follow-up study was designed to identify whether certain criteria assessed during standardized initial screening could predict the response to ovulation induction with clomiphene citrate (CC) in 201 patients presenting with oligomenorrhea or amenorrhea and infertility. Serum FSH levels were within the normal range (1–10 IU/L), and all patients underwent spontaneous or progestin-induced withdrawal bleeding. Initial CC doses were 50 mg daily for 5 days starting on cycle day 3. In the case of an absent response, doses were increased to 100 and 150 mg daily in subsequent cycles. First ovulation with CC was used as the end point. After a complete follow-up (in the case of a nonresponse, at least 3 treatment cycles with daily CC doses up to 150 mg), 156 patients (78%) ovulated. The free androgen index (FAI = testosterone/sex hormone-binding globulin ratio), body mass index (BMI), cycle history (oligomenorrhea vs. amenorrhea), serum androgen (testosterone and/or androstenedione) levels, and mean ovarian volume assessed by transvaginal sonography were all significantly different ($P < 0.01$) in responders from those in nonresponders. FAI was chosen to be the best predictor in univariate analysis. The area under the receiver operating characteristics curve in a multivariate prediction model including FAI, BMI, cycle history, and mean ovarian volume was 0.82.

Patients whose ovaries are less likely to respond to stimulation by FSH due to CC treatment can be predicted on the basis of initial screening characteristics, such as FAI, BMI, cycle history (oligomenorrhea or amenorrhea), and mean ovarian volume. These observations may add to ongoing discussion regarding etiological factors involved in ovarian dysfunction in these patients and classification of normogonadotropic anovulatory infertile women.

INTRODUCTION

Chronic anovulation is a frequent cause of infertility, and approximately 80% of these patients present with serum FSH and estradiol levels within the normal range (WHO group 2) (Rowe *et al.*, 1993) The antiestrogen clomiphene citrate (CC) is considered to be a successful treatment strategy in these patients. It has been documented that approximately 70–80% of these women will become ovulatory (Macgregor *et al.*, 1968) (Gorlitsky *et al.*, 1978) (Shepard *et al.*, 1979) (Hammond *et al.*, 1983) (Polson *et al.*, 1989) (Opsahl *et al.*, 1996), whereas 40–50% of ovulatory women will conceive (Gorlitsky *et al.*, 1978) (Hammond *et al.*, 1983).

Polycystic ovary syndrome (PCOS), usually referred to as chronic hyperandrogenic anovulation, represents a distinct proportion of WHO group 2 anovulatory patients. It is uncertain to what extent PCOS patients are particularly prone to remain resistant to CC medication (Lobo *et al.*, 1982) (Adashi, 1996). Discussion continues regarding the validity of criteria used to diagnose PCOS (Franks, 1995) as well as its relevance for clinical practice. We have previously demonstrated that a distinct overlap exists between endocrine and ultrasound features used by various researchers (van Santbrink *et al.*, 1997). If strict criteria are used for PCOS diagnosis, a large heterogeneous group of normoestrogenic patients will remain unclassified.

Thirty-five years after its first clinical introduction (Greenblatt *et al.*, 1961), CC still remains the first line treatment strategy in normogonadotropic anovulatory patients. Although rising serum FSH levels due to CC interference with estrogen negative feedback may be held responsible for stimulating follicle growth (Jacobson *et al.*, 1968) (Miyake *et al.*, 1983) (Kerin *et al.*, 1985), other mechanisms of action have also been proposed (Adashi, 1996) (Butzow *et al.*, 1995). A significant proportion of treated women, however, do not respond. The aim of this study was to investigate whether clinical, endocrine, and sonographic characteristics during initial screening of normogonadotropic anovulatory infertile women may predict the ovarian response to CC medication. This approach may help to define conditions that prevent the ovary from responding to stimulation by increased FSH levels and to further classify WHO group 2 anovulatory patients.

SUBJECTS AND METHODS

Subjects and study design

Approval for this study was obtained from the human subject committee of the Dijkzigt Hospital/Erasmus University. Between February 1993 and September 1996, 201 patients presenting with oligomenorrhea (interval between vaginal bleeding >35 days and <6 months) or amenorrhea (bleeding interval >6 months) and infertility were recruited. Informed consent was obtained from all participants. All subjects were referred directly by their general practitioner to our infertility unit. None had received

previous ovulation induction medication. Additional inclusion criteria were serum FSH levels within normal limits (1–10 IU/L) (Rowe *et al.*, 1993) (van Santbrink *et al.*, 1995), spontaneous menses or a positive bleeding response to progestagen withdrawal, normal serum PRL and THS levels, body mass index (BMI; weight divided by the square of the patients height) more than 18, and age less than 40 yr.

Clinical, endocrine, and sonographic screening was carried out before initiation of CC therapy. Clinical screening included infertility and cycle history, BMI, previous medication, and/or surgery. Endocrine screening included serum assays of FSH, PRL, TSH, LH, estradiol, androstenedione (AD), testosterone (T), sex hormone-binding globulin (SHBG), cortisol, and dehydroepiandrosterone sulfate. Fasting blood samples were taken randomly between 0800–1000 h before the initiation of therapy. Venous blood samples were centrifuged within 2 h after withdrawal and were stored at -20 C until assayed. Transvaginal sonographic screening included assessment of the ovarian stroma echogenicity (arbitrarily classified from 1–3), ovarian volume (milliliters), and total number of follicles (both ovaries), as described previously (van Santbrink *et al.*, 1997) (Pache *et al.*, 1992b). Serum LH and FSH levels were measured by immunoradiometric assay (Medgenix, Fleurus, Belgium), and T, AD, SHBG, and dehydroepiandrosterone sulfate were determined using RIA kits (Diagnostic Products Corp., Los Angeles, CA), as described previously (Fauser *et al.*, 1991) (van Dessel *et al.*, 1996).

The treatment schedule and assessment of ovarian response were as follows. CC medication was initiated on day 3 after spontaneous or progestagen-induced withdrawal bleeding. The starting dose was 50 mg/day, orally, for 5 subsequent days. In the case of an absent response, daily doses were increased by 50 mg in the next cycle to a maximum dose of 150 mg/day in the following cycle. If ovulation occurred, the dose remained unaltered during subsequent cycles. First ovulation was used as the end point. The duration of follow-up for all patients included in the study was at least three treatment cycles. Ovulation was assessed by midluteal serum progesterone measurement (levels >25 nmol/L indicating ovulation) combined with transvaginal sonographic monitoring of follicle growth until the appearance of a preovulatory follicle (mean diameter, 18 mm) and subsequent follicle rupture, or by biphasic basal body temperature charts. Responders were defined as patients who ovulated during CC therapy, independent of the dose administered. The number of treatment cycles and the CC dose in which first ovulation occurred were recorded. Clomiphene-resistant anovulation (CRA) was defined as patients who do not ovulate despite receiving maximum treatment doses of 150 mg/day.

Data analysis

Distribution of characteristics in patient groups is presented as the mean ± SD. We used the Mann-Whitney U test and the Wilcoxon rank sum W test for exploratory comparison of initial parameters between responders and nonresponders. The univariate and multivariate relation with response to CC was assessed using logistic

regression analysis. The following parameters were used in the analysis: BMI, free androgen index (FAI; T x 100/SHBG), serum T and/or AD concentrations, serum LH levels, cycle history (oligomenorrhea or amenorrhea), cycle duration in case of oligomenorrhea, mean ovarian volume, and follicle number. Backward stepwise elimination was used for the multivariate logistic analysis of prediction of patients being CRA, and P 0.10 was used as a cut-off level for elimination of nonsignificant predictors from the prognostic model. The area under the receiver operating characteristics (ROC) curve (AUC) was used to assess the discriminative ability of the logistic models. The AUC gives the proportion of all pairs of patients (each pair consisting of one patient without and the other patient with a response to CC) in which the model predicts a higher probability of no response for the patient without response. The Statistical Analysis System program (SAS Institute, Cary, NC) was employed for data analysis.

Because selection and estimation were performed for 8 potential predictors on a dataset with only 45 events (CRAs), correction for overfitting was performed (Harrell *et al.*, 1996). The internal validity of the prognostic model was tested by a bootstrapping method in which the selection and estimation process was repeated 200 times. Each of these repetitions consisted of creating a new dataset (bootstrap sample) by drawing cases with replacement from the original data. The backward stepwise elimination process was performed on this dataset, yielding a set of selected predictors and parameter estimates (Harrell *et al.*, 1996). Resulting model estimates of each bootstrap sample were evaluated on the original data, and a shrinkage factor was estimated to correct for statistical overoptimism (Van Houwelingen and Le Cessie, 1990). The Hosmer-Lemeshow goodness of fit test (Hosmer and Lemeshow, 1989) has been used to check for lack of fit of the final model. The logistic coefficients that were corrected by the shrinkage factor have been translated into an easy to use score chart. The scores were calculated by multiplying the shrunk coefficients by 10 and rounding them off to the nearest whole number.

RESULTS

The number of patients who did or did not ovulate after CC medication in increasing doses of 50, 100, and 150 mg daily are depicted in **Figure 1**. Forty-five patients (22.5% of the overall study group) remaining anovulatory were considered having CRA. A total of 432 cycles were analyzed.

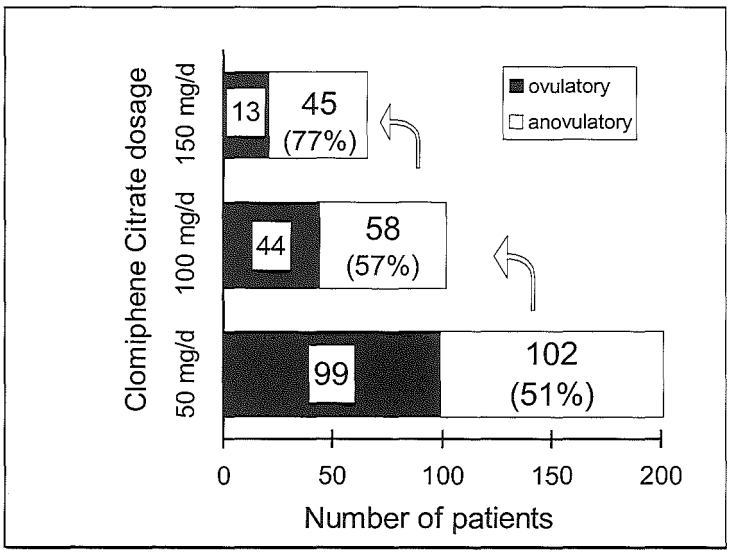


Figure 1: Distribution of normogonadotropic oligomenorrheic or amenorrheic infertile women who do or do not ovulate after CC induction of ovulation in incremental daily doses of 50, 100, or 150 mg for 5 subsequent days. A total of 45 women (22.5% of the overall study group) remain anovulatory.

From the total study group of 201 women, 91 (46%) were considered obese (BMI >26), 101 patients (50%) presented with an elevated FAI (>4.5), 85 patients (42%) presented with hyperandrogenemia (T 3.2 nmol/L and/or AD 16.3 nmol/L) (van Santbrink *et al.*, 1997) (Fauser *et al.*, 1991), and in 125 patients (66%) polycystic ovaries (mean ovarian volume 10.8 mL and/or mean follicle number per ovary 10) (van Santbrink *et al.*, 1997) (Pache *et al.*, 1992b) were diagnosed. Finally, 105 patients (54%) presented with elevated LH (7.0 IU/L) serum levels (van Santbrink *et al.*, 1997) (Fauser *et al.*, 1991).

Table 1: Clinical, endocrine, and ultrasound characteristics (mean ± SD) during initial screening of 201 normogonadotropic oligomenorrheic or amenorrheic infertile women, and separated for patients who do (responders) or do not ovulate (CRA) after CC induction of ovulation.

Screening parameters	Overall group (n = 201)	CC responder (n = 156; 77.5%)	CRA (n = 45; 22.5%)	P value[a]
Clinical				
Age (yr)	28 ± 4.4	28 ± 4.5	27.5 ± 4.5	NS
Primary infertility (n)	145 (72%)	110 (71%)	35 (78%)	NS
Amenorrhea (n)	39 (19%)	22 (14%)	17 (38%)	0.0004
Bleeding interval (days, in case of oligomenorrhea)	79 ± 62	70 ± 56	113 ± 72	<0.0001
BMI (kg/m2)	26.6 ± 6.2	25.5 ± 5.8	30.0 ± 6.6	0.0001
Endocrine				
T (nmol/L)	2.3 ± 0.9	2.1 ± 0.9	2.7 ± 1.0	0.001
AD (nmol/L)	16.5 ± 7.8	15.3 ± 6.5	20.5 ± 10.2	0.001
SHBG (nmol/L)	53 ± 31.7	57 ± 32.5	38.3 ± 23.8	<0.0001
FAI (T x 100/SHBG)	5.9 ± 4.3	4.9 ± 3.4	9.3 ± 5.3	<0.0001
LH (IU/L)	7.8 ± 4.3	7.8 ± 4.3	8.2 ± 4.6	NS
FSH (IU/L)	4.4 ± 1.4	4.5 ± 1.4	4.2 ± 1.4	NS
E2 (pmol/L)	282 ± 233	296 ± 195	234 ± 78	NS
DHEAS (μmol/L)	7.9 ± 3.8	7.9 ± 3.7	7.9 ± 4	NS
TVS				
Total stroma score[b]	3.0 ± 1.0	2.8 ± 1.3	3.3 ± 1.1	0.006
Mean ovarian vol (mL)	10.0 ± 4.4	9.2 ± 5.7	12.2 ± 5.8	0.0007
Mean follicle no.	11.5 ± 5	11 ± 6	12 ± 5	NS

[a] Comparison of CC responders *vs.* CRA (by Mann-Whitney U test).
[b] Arbitrarily defined as one to three per ovary (both ovaries added).

In **Table 1**, clinical, endocrine, and ultrasound characteristics are presented for the overall study group and separately for patients who did or did not ovulate after CC medication. Forty-four percent of patients presenting with amenorrhea (17 of 39) were considered to have CRA, whereas only 17% (28 of 162) of patients with oligomenorrhea showed no response.

Statistical significance in univariate analysis with logistic regression analyses and ROC AUC of the initial parameters are depicted in **Table 2**. The AUCs for FAI and BMI were the highest (0.76 and 0.70, respectively). The ROC curve with the best performance (FAI) and that with the poorest performance (serum LH) are shown in **Figure 2**.

Table 2: Univariate and multivariate logistic regression analyses with score test and area under the ROC curve (AUC) of initial clinical, endocrine, and sonographic screening parameters in 201 normogonadotropic oligomenorrheic or amenorrheic infertile women for the prediction of patients remaining anovulatory after CC induction of ovulation.

Parameters	*P* value	AUC[a]
Univariate analyses		
FAI (T x 100/SHBG)	<0.0001	0.76
BMI (kg/m^2)	<0.0001	0.70
Mean ovarian vol	0.0001	0.67
Hyperandrogenemia (elevated T and/or AD)	0.0007	0.64
Oligomenorrhea or amenorrhea	0.0005	0.62
Mean follicle no.	0.1	0.58
Bleeding interval in case of oligomenorrhea	0.42	0.53
LH (IU/L)	0.5	0.52
Multivariate analysis		
Prediction model for CRA[b]		0.82

[a] Area under the ROC curve.

[b] Combination of four initial screening parameters: FAI, BMI, cycle history (oligomenorrhea or amenorrhea), and mean ovarian volume.

Of the 201 patients, 187 had complete data on the variables used in the multivariate analysis. Using the backward elimination procedure, 4 variables were selected in the final model: 1) FAI, 2) BMI, 3) cycle history (oligomenorrhea or amenorrhea), and 4) mean ovarian volume. By using the combined information of these 4 variables, the AUC further improved to 0.82 (**Table 2** and **Figure 2**).

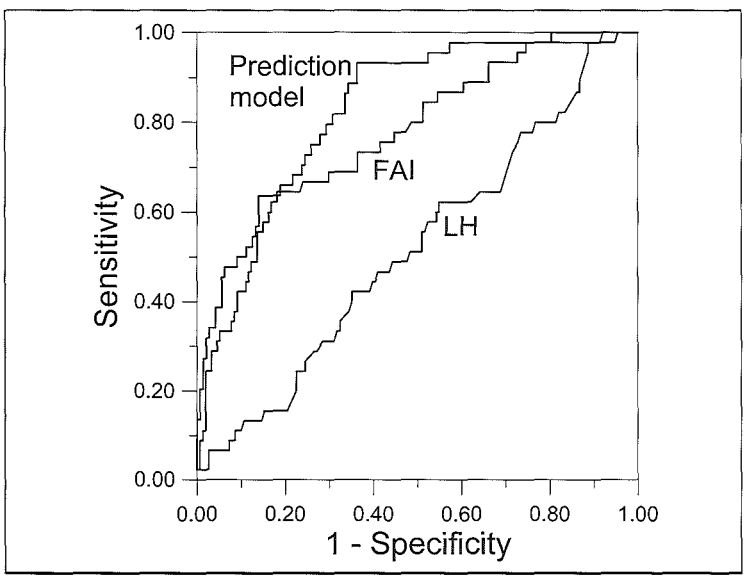


Figure 2: ROC curve of serum LH concentration, FAI, or the prediction model (FAI, BMI, cycle history, and mean ovarian volume combined) for predicting CRA in a total group of 201 normogonadotropic oligomenorrheic or amenorrheic infertile women.

The bootstrap procedure revealed that these 4 predictors were selected in over two thirds of the bootstrap samples, whereas all other candidate variables were selected in less than half of the samples, which illustrates the stability of the final model. The shrinkage factor was estimated from the bootstrap procedure to be 0.82, indicating that when this study is replicated many times, the resulting coefficients of the final multivariate model are, on the average, 18% smaller. This was incorporated in the calculation of the scores.

The scores for different parameters are depicted in **Table 3**, and resulting probability scores for patients remaining anovulatory after CC medication are shown

Table 3 Score chart with the four initial screening parameters of the final model for prediction of patients remaining anovulatory after CC induction of ovulation in normogonadotropic oligomenorrheic or amenorrheic infertile women (total scores, 0–53)

Initial screening parameters	Score[a]
FAI (T x 100/SHBG)	
<2	0
2–3	1
3–4	2
4–5	3
5–6	5
6–8	6
8–11	10
>11	14
BMI (kg/m^2)	
<20	0
20–21.5	2
21.5–23	3
23–25	4
25–27	6
27–31	8
31–35	12
>35	15
Mean ovarian vol (mL)	
<6	0
6–7	2
7–8	2
8–9	3
9–11	5
11–13	6
13–16	9
>16	11
Cycle history	
Oligomenorrhea	0
Amenorrhea	13
Total score	—

[a] Encircle the scores related to each category of the screening parameters and add them together. Correspond the total score to the score chart (Fig. 3). As an example, a new amenorrheic patient had the following findings: FAI = 8.7, BMI = 29.4, and mean ovarian volume = 13 ml. Scores are 10 for FAI, 8 for BMI, 9 for mean ovarian volume, and 13 for having amenorrhea. The total score is 40, and the corresponding probability to be CRA is 65%.

in **Figure 3**. The Hosmer-Lemeshow goodness of fit test showed no lack of fit of the final model to the data (P = 0.3).

DISCUSSION

It may be helpful for further classification of normogonadotropic anovulatory infertility and for evaluating pathophysiological factors involved in ovarian abnormalities in these patients to study in a longitudinal fashion whether initial screening parameters may predict success, failure, or complications of induction of ovulation. The present study was designed to investigate as a first step whether ovarian response after CC medication could be predicted. It is established in the literature that approximately 75% of patients will ovulate, and less than 50% of the total population will conceive after CC as first line medication. If patients remaining anovulatory despite CC therapy could be identified beforehand, ineffective and time-consuming CC treatment could be prevented. This may be helpful, particularly for women of advanced reproductive age. Further studies, however, are required to investigate whether alternative primary treatment options are cost effective. Moreover, etiological factors involved in ovarian dysfunction could be identified in this heterogeneous patient group, as subjects whose ovaries will or will not respond to increased FSH stimulation may be differentiated.

For this first analysis of response to CC medication we decided to focus on ovulation rather than conception. Ovulation is biologically relevant and most closely connected to the desired effects of CC medication. Analysis of conception as the end

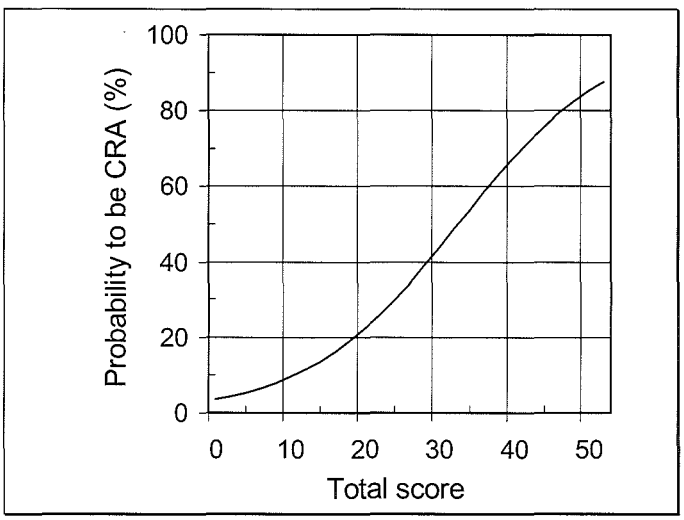


Figure 3 Total score from the score chart and the corresponding probability for a given woman to remain anovulatory after CC induction of ovulation.

point requires a comprehensive study of other potential confounders, such as tubal factor, sperm factor, and endometrial function. Four initial screening parameters (FAI, BMI, mean ovarian volume, and cycle history) could be identified, predicting patients remaining anovulatory after CC medication. A combination of these parameters showed good predictive power, with a ROC AUC of 0.82. Several studies have been published recently regarding the use of a similar multivariate model for predicting chances for conception in infertile patients with regular cycles (Eimers *et al.*, 1994) (Collins *et al.*, 1995) (Snick *et al.*, 1997). Various researchers have investigated the predictive value of clinical and endocrine screening parameters for the response to CC. Only a positive correlation between body weight and the dose of CC required to induce ovulation has been established (Lobo *et al.*, 1982) (Shoham *et al.*, 1990). A recent study has indicated that increased BMI is the only initial parameter that is significantly different between responders and nonresponders (Kousta *et al.*, 1997). To our knowledge this is the first time a multivariate prediction model has been applied in the treatment of anovulatory infertility. We could demonstrate that patients suffering from amenorrhea, obesity, increased ovarian volume, and elevated androgen levels (a complex of signs, symptoms, and ultrasound and endocrine findings frequently referred to as PCOS) are most likely to remain anovulatory after CC induction of ovulation. A model can be used to predict chances for an individual patient to remain anovulatory by calculating a total score on the basis of these initial screening characteristics. Further studies should validate the prediction model in a new group of patients. The present study also suggests that LH concentrations do not predict ovarian response after CC medication in accordance with recent observations by others (Kousta *et al.*, 1997). These data oppose the concept that elevated LH is implemented in ovarian dysfunction in these patients (Yen, 1980). However, the assessment of LH levels in anovulatory patients is problematic due to effects of timing, the immunoassays used, and the pulsatile nature of LH release (Fauser and De Jong, 1993).

Together, these observations suggest that obese hyperandrogenic women are less likely to respond to increased stimulation by FSH, suggesting that these factors are instrumental in follicle maturation arrest (Franks *et al.*, 1998) and an increased FSH threshold (Polson *et al.*, 1989). The correlation between BMI and ovarian response after CC treatment suggests that much emphasis should be focused on weight reduction. However, it should be realized that scientific proof for this approach is lacking, and that weight reduction may not necessarily result in normal response. Previous work from our group demonstrated indeed that long term androgen medication in female to male transsexuals induces polycystic ovary morphology, characterized by increased ovarian size, augmented follicle number, and stroma hyperplasia (Pache *et al.*, 1991). Assessment of steroid levels in follicle fluid obtained from polycystic ovaries suggested that disturbed dominant follicle selection in hyperandrogenic patients may result from disrupted enhancement of FSH-induced aromatase activity (Pache *et al.*, 1992a). This could be due to intraovarian

dysregulation of FSH action, which precludes a normal response (i.e. follicle growth and ovulation) after incremental FSH levels elicited by CC medication. Factors involved may include locally produced growth factors or insulin resistance (Giudice, 1992) (Nestler, 1997). Alternative explanations for a nonresponse after CC may include 1) an abnormal hypothalamic/pituitary response to steroid feedback resulting in an insufficient rise in FSH after CC, 2) individual differences in the FSH isohormone profile resulting in discrepancies in bioactive FSH concentrations despite similar immunoreactive FSH levels, 3) additional, as yet unidentified, mechanisms responsible for at least part of CC actions. Recently reported CC-induced changes in the insulin-like growth factor system (Butzow *et al.*, 1995) may be relevant in this regard. It may be speculated that improved insight into any of these factors (intraovarian dysregulation (Fauser and Van Heusden, 1997), hypothalamic/pituitary dysfunction (Berga and Daniels, 1997), FSH heterogeneity (Ulloa-Aguirre *et al.*, 1995), or the insulin-like growth factor system (Giudice, 1992)) may eventually result in additional predictors of the CC response. Further studies focusing on insulin resistance may also be of interest, as recent reports suggest that ovarian dysfunction may improve after the use of insulin-sensitizing agents (Sattar *et al.*, 1998). Preliminary observations (Imani, B. and Fauser, B. C. J. M., unpublished observations) suggest that fasting insulin levels are increased in CRA patients. It requires further study to clarify whether insulin resistance is a determining factor in an abnormal response to FSH independent from androgen concentrations, as the majority of studies propose that insulin resistance is associated with PCOS through increased thecal cell androgen production (Nestler, 1997).

In conclusion, this study demonstrates that it is possible to predict patients remaining anovulatory during CC induction of ovulation using criteria that are directly associated with PCOS, predominantly obesity and hyperandrogenemia. Further studies should establish whether the occurrence of pregnancies after CC medication can also be predicted and whether similar factors are involved. The identification of initial characteristics that predict the ovarian response to ovulation induction therapy may help to further classify the heterogeneous group of normogonadotropic anovulatory infertile women. The present study suggests that hyperandrogenemia and obesity are crucial in inducing ovarian abnormalities that are less likely to respond to increased stimulation by FSH.

Footnote

1 This work was supported by the Stichting Voortplantingsgeneeskunde Rotterdam and a research grant from Ferring Nederland.

References

Adashi EY. Ovulation induction: clomiphene citrate. In: Adashi EY, Rock JA, Rosenwaks Z, editors. *Reproductive endocrinology, surgery, and technology*. Philadelphia: Lippencott-Raven, 1996:1182-1206.

Berga SL, Daniels TL. (1997) Can polycystic ovary syndrome exist without concomitant hypothalamic dysfunction? *Semin Reprod Endocrinol.* 15, 169-75.

Butzow TL, Kettel LM, Yen SS. (1995) Clomiphene citrate reduces serum insulin-like growth factor I and increases sex hormone-binding globulin levels in women with polycystic ovary syndrome. *Fertil Steril.* 63, 1200-3.

Collins JA, Burrows EA, Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril.* 64, 22-8.

Eimers JM, te Velde ER, Gerritse R, Vogelzang ET, Looman CW, Habbema JD. (1994) The prediction of the chance to conceive in subfertile couples. *Fertil Steril.* 61, 44-52.

Fauser BC, Pache TD, Lamberts SW, Hop WC, de Jong FH, Dahl KD. (1991) Serum bioactive and immunoreactive luteinizing hormone and follicle- stimulating hormone levels in women with cycle abnormalities, with or without polycystic ovarian disease. *J Clin Endocrinol Metab.* 73, 811-7.

Fauser BC, De Jong FH. (1993) Gonadotropins in polycystic ovarian syndrome. *Ann N Y Acad Sci.* 687, 150-61.

Fauser BC, Van Heusden AM. (1997) Manipulation of human ovarian function: physiological concepts and clinical consequences. *Endocr Rev.* 18, 71-106.

Franks S. (1995) Polycystic ovary syndrome. *N Engl J Med.* 333, 853-861.

Franks S, Mason H, Willis D. Disorders of folliculogenesis in polycystic ovary syndrome. In: Filicori M, Flamigni C, editors. *Ovulation induction update '98*. London: The parthenon publishing group, 1998:37-40.

Giudice LC. (1992) Insulin-like growth factors and ovarian follicular development. *Endocr Rev.* 13, 641-69.

Gorlitsky GA, Kase NG, Speroff L. (1978) Ovulation and pregnancy rates with clomiphene citrate. *Obstet Gynecol.* 51, 265-9.

Greenblatt RB, Barfield WE, Jungck EC, Ray AW. (1961) Induction of ovulation with MRL/41: preliminary report. *J Am Med Assoc.* 178, 101-4.

Hammond MG, Halme JK, Talbert LM. (1983) Factors affecting the pregnancy rate in clomiphene citrate induction of ovulation. *Obstet Gynecol.* 62, 196-202.

Harrell FE, Jr., Lee KL, Mark DB. (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 15, 361-87.

Hosmer DW, Lemeshow S. Applied logistic regression. New York, NY: John Wiley & Sons Inc, 1989:140-145.

Jacobson A, Marshall JR, Ross GT, Cargille CM. (1968) Plasma gonadotropins during clomiphene induced ovulatory cycles. *Am J Obstet Gynecol.* 102, 284-90.

Kerin JF, Liu JH, Phillipou G, Yen SS. (1985) Evidence for a hypothalamic site of action of clomiphene citrate in women. *J Clin Endocrinol Metab.* 61, 265-8.

Kousta E, White DM, Franks S. (1997) Modern use of clomiphene citrate in induction of ovulation. *Hum Reprod Update.* 3, 359-65.

Lobo RA, Gysler M, March CM, Goebelsmann U, Mishell Jr DR. (1982) Clinical and laboratory predictors of clomiphene response. *Fertil Steril.* 37, 168-174.

Macgregor AH, Johnson JE, Bunde CA. (1968) Further clinical experience with clomiphene citrate. *Fertil Steril.* 19, 616-22.

Miyake A, Tasaka K, Sakumoto T, Kawamura Y, Nagahara Y, Aono T. (1983) Clomiphene citrate induces luteinizing hormone release through hypothalamic luteinizing hormone-releasing hormone in vitro. *Acta Endocrinol (Copenh).* 103, 289-92.

Nestler JE. (1997) Role of hyperinsulinemia in the pathogenesis of the polycystic ovary syndrome, and its clinical implications. *Semin Reprod Endocrinol.* 15, 111-22.

Opsahl MS, Robins ED, O'Connor DM, Scott RT, Fritz MA. (1996) Characteristics of gonadotropin response, follicular development, and endometrial growth and maturation across consecutive cycles of clomiphene citrate treatment. *Fertil Steril.* 66, 533-9.

Pache TD, Chadha S, Gooren LJ, Hop WC, Jaarsma KW, Dommerholt HB, et al. (1991) Ovarian morphology in long-term androgen-treated female to male transsexuals. A human model for the study of polycystic ovarian syndrome? *Histopathology.* 19, 445-52.

Pache TD, Hop WC, de Jong FH, Leerentveld RA, van Geldorp H, Van de Kamp TM, et al. (1992a) 17 beta-Oestradiol, androstenedione and inhibin levels in fluid from individual follicles of normal and polycystic ovaries, and in ovaries from androgen treated female to male transsexuals. *Clin Endocrinol (Oxf).* 36, 565-71.

Pache TD, Wladimiroff JW, Hop WC, Fauser BC. (1992b) How to discriminate between normal and polycystic ovaries: transvaginal US study. *Radiology.* 183, 421-3.

Polson DW, Kiddy DS, Mason HD, Franks S. (1989) Induction of ovulation with clomiphene citrate in women with polycystic ovary syndrome: the difference between responders and nonresponders. *Fertil Steril.* 51, 30-4.

Rowe PJ, Comhaire FH, Hargreave TB, Mellows HJ. (1993) *WHO manual for the standardized investigation and diagnosis of the infertile couple.*, Cambridge University Press,Cambridge, UK.

Sattar N, Hopkinson ZE, Greer IA. (1998) Insulin-sensitising agents in polycystic-ovary syndrome. *Lancet.* 351, 305-7.

Shepard MK, Balmaceda JP, Leija CG. (1979) Relationship of weight to successful induction of ovulation with clomiphene citrate. *Fertil Steril.* 32, 641-5.

Shoham Z, Borenstein R, Lunenfeld B, Pariente C. (1990) Hormonal profiles following clomiphene citrate therapy in conception and nonconception cycles. *Clin Endocrinol (Oxf).* 33, 271-8.

Snick HK, Snick TS, Evers JL, Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod.* 12, 1582-8.

Ulloa-Aguirre A, Midgley AR, Jr., Beitins IZ, Padmanabhan V. (1995) Follicle-stimulating isohormones: characterization and physiological relevance. *Endocr Rev.* 16, 765-87.

van Dessel HJ, Schoot BC, Schipper I, Dahl KD, Fauser BC. (1996) Circulating immunoreactive and bioactive follicle stimulating hormone concentrations in anovulatory infertile women and during gonadotrophin induction of ovulation using a decremental dose regimen. *Hum Reprod.* 11, 478-85.

van Houwelingen JC, Le Cessie S. (1990) Predictive value of statistical models. *Stat Med.* 9, 1303-25.

van Santbrink EJ, Hop WC, van Dessel TJ, de Jong FH, Fauser BC. (1995) Decremental follicle-stimulating hormone and dominant follicle development during the normal menstrual cycle. *Fertil Steril.* 64, 37-43.

van Santbrink EJ, Hop WC, Fauser BC. (1997) Classification of normogonadotropic infertility: polycystic ovaries diagnosed by ultrasound versus endocrine characteristics of polycystic ovary syndrome. *Fertil Steril.* 67, 452-8.

Yen SS. (1980) The polycystic ovary syndrome. *Clin Endocrinol (Oxf).* 12, 177-207.

8

Predictors of chances to conceive in ovulatory patients during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility

ABSTRACT

The present prospective follow-up study was designed to identify whether clinical, endocrine, or ultrasound characteristics assessed by standardized initial screening of normogonadotropic oligo/amenorrheic infertile patients could predict conception in 160 women who reached ovulation after clomiphene citrate (CC) medication. Additional inclusion criteria were total motile sperm count of the partner above 1 million and a negative history for any tubal disease. Daily CC doses of 50 mg (increasing up to 150 mg in case of absent ovarian response) from cycle days 3–7 were used. First conception (defined as a positive urinary pregnancy test) was the end point for this study. A cumulative conception rate of 73% was reached within 9 CC-induced ovulatory cycles. Patients who did conceive presented more frequently with lower age ($P < 0.0001$) and amenorrhea ($P < 0.05$) upon initial screening. In a univariate analysis, patients with elevated initial serum LH concentrations (>7.0 IU/L) had a higher probability of conceiving ($P < 0.01$). In a multivariate analysis, age and cycle history (oligomenorrhea vs. amenorrhea) were identified as the only significant parameters for prediction of conception.

These observations suggest that there is more to be gained from CC ovulation induction in younger women presenting with profound oligomenorrhea or amenorrhea. Screening characteristics involved in the prediction of ovulation after CC medication in normogonadotropic oligo/amenorrheic patients (body weight and hyperandrogenemia, as shown previously) are distinctly different from predictors of conception in ovulatory CC patients (age and the severity of cycle abnormality). This disparity suggests that the FSH threshold (magnitude of FSH required for stimulation of ongoing follicle growth and ovulation) and oocyte quality (chances for conception in ovulatory cycles) may be differentially regulated.

INTRODUCTION

The synthetic antiestrogen clomiphene citrate (CC) represents an easy to use, convenient, inexpensive, and safe first choice medication in normogonadotropic oligo/amenorrheic infertility (WHO group 2) (Rowe *et al.*, 1993). Life-table analysis of pregnancy rates after CC medication and prediction of treatment outcome have been the subject of extensive investigation (Macgregor *et al.*, 1968) (Gorlitsky *et al.*, 1978) (Shepard *et al.*, 1979) (Lobo *et al.*, 1982) (Hammond *et al.*, 1983) (Polson *et al.*, 1989) (Shoham *et al.*, 1990) (Opsahl *et al.*, 1996) (Kousta *et al.*, 1997). Cumulative pregnancy rates after CC treatment between 37–97% have been reported (Gorlitsky *et al.*, 1978) (Shepard *et al.*, 1979) (Hammond *et al.*, 1983). A positive correlation was established between body weight and the CC dose required to induce ovulation (Lobo *et al.*, 1982) (Kousta *et al.*, 1997). Moreover, recent studies have indicated that body mass index (BMI) is significantly higher in nonresponders (Polson *et al.*, 1989) (Kousta *et al.*, 1997). Limited information is available, however, concerning the predictive value of initial screening characteristics for treatment outcome (Adashi, 1996), and previous investigators were unable to identify predictors for conception after CC induction of ovulation (Hammond *et al.*, 1983) (Shoham *et al.*, 1990) (Kousta *et al.*, 1997). All the above-mentioned studies focused on the prediction of treatment outcome in the entire group of infertile patients who started with CC medication. In contrast, we focused separately on the prediction of ovulation after CC administration (Imani *et al.*, 1998) and the prediction of conception in ovulatory CC-treated women. This approach seems more appropriate, because statistical bias due to selective drop out from the study for reasons of persistent anovulation despite increasing doses of CC medication is eliminated.

Our group could recently establish that obese hyperandrogenic amenorrheic patients are more likely to be resistant to CC medication (Imani *et al.*, 1998). We now report on initial clinical, endocrine, and sonographic screening characteristics of normogonadotropic oligomenorrheic or amenorrheic infertile women achieving ovulatory cycles after CC medication in an attempt to identify factors predicting chances for conception in these patients. The separate focus on the prediction of conception and ovulation after CC treatment may allow to differentiate among factors affecting oocyte quality independently from follicle development.

SUBJECTS AND METHODS

Subjects and study protocol

Between February 1993 and December 1997, 160 couples presenting with oligomenorrhea (bleeding intervals between 35 days and 6 months) or amenorrhea (bleeding interval >6 months) and infertility attending our unit were included in the present study. Additional inclusion criteria were 1) serum FSH levels within normal

limits (1–10 IU/L) (Rowe *et al.*, 1993) (van Santbrink *et al.*, 1997) (Schipper *et al.*, 1998) and normal serum PRL and TSH levels, 2) spontaneous menses or positive bleeding response to progestagen withdrawal, 3) ovulatory cycles after CC induction of ovulation, 4) BMI (weight divided by height squared) greater than 18, 5) age between 19–40 years, 6) a total motile sperm count [TMC = ejaculate volume (milliliters) × sperm concentration (10^6 / mL) × percentage of progressive motile sperm] of the partner above 1 million (Ombelet *et al.*, 1997), 7) negative history for any tubal pathology, and 8) no indication for intrauterine insemination. Study approval was obtained from the human subject committee of the Dijkzigt Hospital/Erasmus University, and informed consent was obtained from all subjects included. A standardized clinical, endocrine, and sonographic screening took place before initiation of induction of ovulation with CC medication.

Ovulation after CC treatment was assessed by midluteal serum progesterone (P) levels above 25 nmol/L, combined with transvaginal sonographic monitoring of follicle growth until visualization of a preovulatory follicle (mean diameter, >18 mm) and subsequent disappearance or biphasic basal body temperature charts, as described previously (Imani et al., 1998) Clomiphene citrate was administered at a daily dose of 50 mg (increased to 100 and 150 mg in subsequent cycles in the case of absent ovarian response) from cycle days 3–7 after initiation of spontaneous or progestin-induced withdrawal bleeding. Conception was defined as a positive urinary pregnancy test (Clearview, hCG II, Unipath Ltd., Bedford, UK) more than 3 days after the expected menses, and ongoing pregnancy was defined as sonographic assessment of an intrauterine gestational sac with positive heart beat.

History and clinical screening included assessment of duration of infertility, whether infertility was primary or secondary, cycle history, previous medication and/or surgery, and BMI. Endocrine screening included serum assays of FSH, LH, estradiol (E_2), testosterone (T), androstenedione (AD), sex hormone-binding globulin (SHBG), and P. Fasting venous blood samples were taken on a random day between 8:00–10:00 A.M., as indicated previously (Imani *et al.*, 1998). Blood samples were centrifuged within 2 hours after withdrawal and were stored at –20 C until assayed. Serum LH and FSH levels were measured by immunofluorometric assay (Amerlite, Ortho-Clinical Diagnostics, Amersham, Aylesbury, U.K.), as described previously (Schipper *et al.*, 1998). P levels were measured by radioimmunoassay (RIA), as described previously (de Jong *et al.*, 1974). Serum E_2, T, AD, and SHBG levels were estimated using RIA kits provided by Diagnostic Products Corp. (DPC, Los Angeles, CA), as described previously (Fauser *et al.*, 1991). Intra- and interassay coefficients of variation were less than 3% and 8% for FSH, less than 5% and 15% for LH, less than 16% and 17% for P, less than 5% and 7% for E_2, less than 3% and 5% for T, less than 8% and 11% for AD, and less than 4% and 5% for SHBG, respectively. Transvaginal pelvic ultrasound (Model EUB-415, Hitachi Medical Corp., Tokyo, Japan) was performed by a single observer (B.I.) and included the assessment of ovarian stroma amount and echogenicity (arbitrarily classified from one to three per ovary), ovarian volume

(milliliters), and total number of follicles (both ovaries), as described previously (van Santbrink *et al.*, 1997) (Pache *et al.*, 1992). Semen analyses were performed according to WHO guidelines (1992) and comprised ejaculate volume (milliliters), number of spermatozoa per mL (10^6 spermatozoa/mL), percentage of progressive motile spermatozoa, and percentage of normal forms (Organization, 1992).

Data analysis

A *P* value of 0.05 was chosen as the threshold level for statistical significance. Cox regression has been used for life-table analysis of conception rates during CC treatment (Cox, 1972). The number of ovulatory CC treatment cycles was the time variable for multivariate analyses. Censoring was defined as definitive discontinuation of CC therapy without conception or end of follow-up (February 1998). To analyze the effect of the severity of the cycle abnormality on chances to conceive after CC treatment, we arbitrarily divided the cycle histories of the patients into four categories; interval between periods of 5–6 weeks (n = 56), 6–9 weeks (n = 50), 9–26 weeks (n = 25), and greater than 26 weeks (*i.e.* amenorrhea) (n = 29). The univariate relation was assessed between the variables listed in **Table 1** and the time interval between the first ovulation after CC medication and conception using the Kaplan-Meier method (Kaplan and Meier, 1958) for categorical and the Cox regression (Cox, 1972) for continuous variables. The Log-rank test has been used to denote statistical significance in life-table analyses. The multivariate analysis was performed with the method of forward stepwise selection to gain a better insight into the interdependence between initial screening parameters. This method can explain why a variable that was significantly different in univariate analysis was not selected in the final model. The prognostic impact of variables was expressed as a fecundability ratio, which is equivalent to the hazard ratio in survival analysis. For instance, a fecundability ratio of 0.9 for an unfavorable group means that the conception rate per ovulatory CC treatment cycle is 10% lower compared to that in the favorable group. In some couples (n = 25), no sperm analysis was performed because of the short time interval between initiation of CC medication and conception. A statistical imputation technique has been applied using multiple conditional mean imputation to fill in these missing sperm parameters (Rubin, 1978). The value of semen parameters was estimated using the time until conception and patient characteristics related to sperm parameters in the nonmissing group. Data were analyzed using the commercially available software package SPSS, Inc. (Chicago, IL).

RESULTS

A total of 82 women (51% of the ovulatory group) conceived, and 73 (46%) reached an ongoing pregnancy from the total of 160 patients fulfilling the in/exclusion criteria. Sixty-nine were singleton and 4 were twin pregnancies (data not shown). Initial

screening parameters of 9 patients who had a miscarriage after CC (11% of overall CC conceptions) were not different from those of the remaining 73 patients who reached an ongoing pregnancy (data not shown). One woman who conceived (twin pregnancy) during her first CC treatment cycle was frequently monitored on an out-patient basis due to abdominal discomfort and enlarged ovaries. There was no case of severe ovarian hyperstimulation syndrome.

The life-table analysis of cumulative conception rates (CCR) of patients who ovulated after CC are indicated in **Figure 1** for the total group, and separately for different dose groups. A cumulative conception rate of 47% was reached within three cycles from first ovulation, and a CCR of 73% was reached within nine CC-induced ovulatory cycles. Patients using daily doses of 50, 100, and 150 mg CC reached

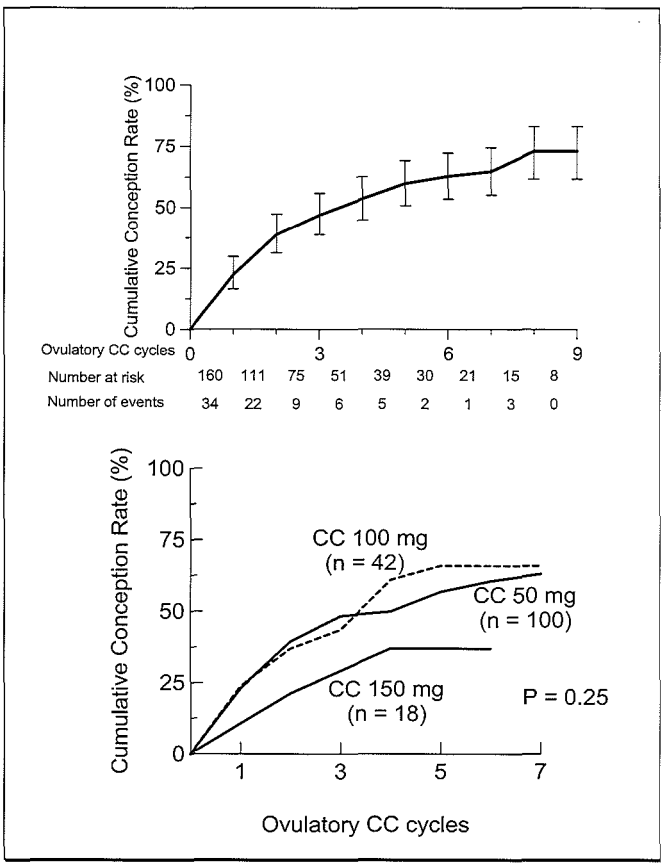


Figure 1: Life table analysis of cumulative conception rates (CCR) in 160 normogonadotropic oligo-amenorrheic infertile patients who ovulated following CC medication. CCR's (including absolute number of patients at risk and number of events (= conceptions)) are presented for the total study group (upper panel) (vertical lines represents 95% Confidence Intervals), and separately for women who reached ovulatory CC cycles with 50, 100, or 150 mg daily for 5 subsequent days (lower panel). N represents the initial number of patients at risk per dose group. P = Log Rank test P value.

cumulative conception rates of 57%, 66%, and 38% within 5 ovulatory cycles, respectively ($P_{\log rank} = 0.25$; **Figure 1**). At higher doses, chances for conception and ongoing pregnancy are not statistically significantly reduced, although absolute CCR (n = 18) were low in the 150-mg CC group. The overall mean duration of follow-up was 4 ± 3 months and 3.2 ± 2.6 ovulatory CC cycles.

Initial screening characteristics of the overall group of patients who ovulated after

Table 1: Initial clinicalscreening characteristics and sperm parameters of partners (median and range) of 160 normogonadotropic oligomenorrheic or amenorrheic infertile women who ovulated after CC induction of ovulation (overall group) and did or did not (CC failure) conceive, endocrine, and ultrasound

Screening parameters		Overall group (n = 160)	Conceived (n = 82; 51%)	CC failure (n = 78; 49%)	*P* value[a]
Clinical					
	Age (yr)	28 ± 4	27 ± 4	29 ± 4	0.0001
	Infertility duration (yr)	1.9 ± 2.3	1.6 ± 1.4	2.2 ± 2.9	0.23
	Primary infertility (n)	116	62	54	0.40
	Amenorrhea (n)	29	18 (62)	11 (38)	0.04
	Bleeding interval in 4 categories				0.05[b]
	5–6 weeks, n (%)	56 (35)	22	34	
	6–9 weeks, n (%)	50 (31)	29	21	
	9–26 weeks, n (%)	25 (16)	13	12	
	>26 weeks, n (%)	29 (18)	18	11	
	BMI (kg/m^2)	26 ± 6	26 ± 6	25 ± 6	0.46
Endocrine					
	LH (IU/L)	7.6 ± 4.2	8.2 ± 4.3	6.9 ± 4.0	0.11
	T (nmol/L)	2.3 ± 1.0	2.3 ± 1.0	2.3 ± 1.0	0.85
	AD (nmol/L)	15.0 ± 6.9	15.8 ± 7.9	14.0 ± 5.7	0.16
	FAI[c]	5.4 ± 3.9	5.5 ± 4.1	5.1 ± 3.7	0.20
	SHBG (nmol/L)	56 ± 31	54 ± 27	58 ± 35	0.24
	E_2 (pmol/L)	259 ± 175	229 ± 145	290 ± 198	0.10
Ultrasound					
	Mean ovarian vol (mL)	9.3 ± 3.7	9.2 ± 3.7	9.4 ± 3.7	0.97
	Mean follicle number	12 ± 5	12 ± 5	12 ± 5	0.48
	Total stroma score[d]	3.2 ± 1.1	3.1 ± 1.1	3.3 ± 1.2	0.66
TMC[e]		73 (1.1–492)	75 (1.5–303)	72 (1.1–492)	0.32[f]
% Normal morphology		24 (1–50)	20 (1–50)	25 (1–50)	0.34[f]

Values are the mean ± SD. Underlined values are statistically significant.

[a] Comparison of CC-conceived *vs.* CC failure (univariate Cox regression).

[b] Chances to conceive differ among four categories of bleeding interval.

[c] FAI = T x 100/SHBG.

[d] Arbitrarily defined as one to three per ovary (both ovaries added).

[e] Total motile sperm count

[f] Computed with multiple imputation of missing values.

CC and separately for those women who conceived (n = 82) *vs.* those who did not (n = 78) are depicted in **Table 1**. Age, the severity of cycle abnormality (oligomenorrhea *vs.* amenorrhea), and cycle duration, arbitrarily classified in four categories (see also *Materials & Methods*), were significantly different in univariate analysis.

Age (cut-off of 30 yr), cycle history (oligomenorrhea *vs.* amenorrhea), and initial

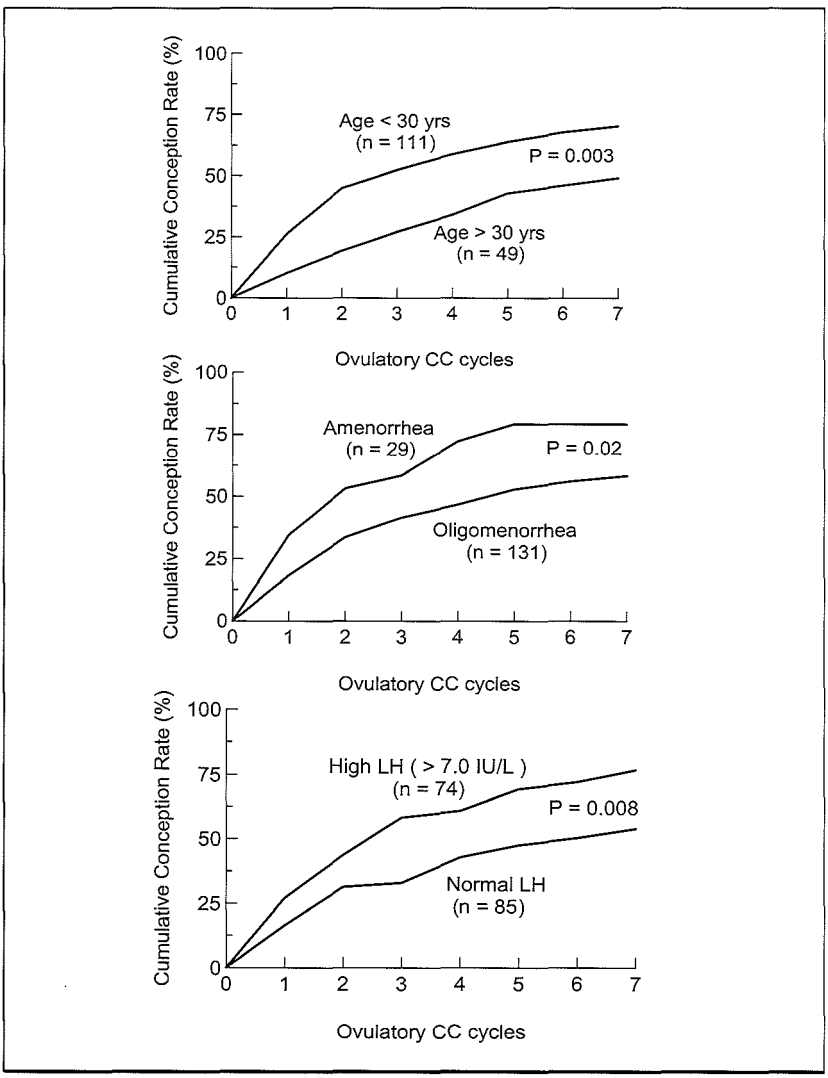


Figure 2 Univariate analysis of cumulative conception rates in 160 normogonadotropic oligoamenorrheic infertile patients per ovulatory CC cycle. Initial screening parameters shown are: 1) Patient's age (cut-off level at 30 years) (upper panel). 2) Cycle history (oligomenorrhea versus amenorrhea) (middle panel). 3) Initial serum LH concentrations (cut-off level of 7.0 IU/L) (lower panel). n represents the initial number of patients at risk. P = Log Rank test P value.

serum LH level (cut-off level of 7.0 IU/L) in univariate analyses for CCR are depicted in **Figure 2**. The percentage of ongoing pregnancies per conception for patients with elevated (initial serum LH level ≥7.0 IU/L) or normal initial serum LH levels were 84.8% and 94.4%, respectively (*P* value for difference in proportion ongoing pregnancies = 0.16, and 95% confidence interval for difference = -3% to 23%). The cut-off value for normal (*i.e.* 7.0 IU/L) was chosen on the basis of a previous study from our group in normoovulatory controls (mean + 1 SD) (van Santbrink *et al.*, 1997).

A total number of 159 patients had complete data on the variables used in the multivariate analyses. Univariate analysis and forward stepwise multivariate analyses of all initial parameters for the prediction of chances to conceive in ovulatory patients treated with CC are depicted in **Table 2**. During the stepwise multivariate analysis for the prediction of chances to conceive, age, and cycle history (amenorrhea *vs.* oligomenorrhea) entered into the model (step 1 and 2, respectively). The multivariate-adjusted fecundability ratio for age was 0.90 (95% confidence interval, 0.85–0.95), and that for amenorrhea 0.54 (95% confidence interval, 0.32–0.93).

Table 2: Forward stepwise multivariate analyses of initial screening characteristics for the prediction of chances to conceive in 159 normogonadotropic oligo-amenorrheic infertile women who ovulated after CC induction of ovulation

	Univariate	Multivariate	
Analyses steps			
	0[a]	1	2
Screening parameters			
Clinical			
Age (yr)	0.0001	In model	In model
Amenorrhea (n = 29)	0.04	0.02	In model
Bleeding interval (in four categories)	0.05	0.06	0.30
Endocrine			
LH (IU/L)	0.11	0.36	0.26
FAI[b]	0.20	0.56	0.38
AD (nmol/L)	0.16	0.38	0.15
E_2 (pmol/L)	0.10	0.26	0.39

Numbers are *P* values for inclusion in the model. *Underlined* numbers are significant at $P < 0.05$.

[a] Only screening parameters with a univariate $P \leq 0.2$ (see **Table 1**) are shown. In the univariate analysis (step 0), three variables reach statistical significance (*underlined*). In step 1 of the multivariate analysis the variable with the highest prognostic information (age) is selected. After the first step, amenorrhea still reaches statistical significance and, therefore, is selected in the second step. Thereafter, no additional variable is statistically significant anymore, indicating that the model cannot be improved by selecting a subsequent parameter.

[b] FAI = T x 100/SHBG.

DISCUSSION

The present prospective follow-up study was designed to evaluate whether initial screening characteristics of 160 normogonadotropic oligo/amenorrheic infertile women could predict conception during ovulatory CC-induced cycles. Although CC medication has been the focus of prevalent research, limited information is available regarding the prediction of conception as treatment outcome (Adashi, 1996). Reported cumulative pregnancy rates vary between 37-97%. Most studies, however, suffer from methodological difficulties and different inclusion/exclusion criteria. For the first time, our group has focused on ovulation and conception in separate steps, taking into account that a significant proportion (23%) of patients who remain anovulatory after CC medication (Imani *et al.*, 1998) have no chance of conception. Inclusion of these patients in a study focusing on conception causes statistical bias. This separate focus may offer a better insight into the potential predictive power of initial screening characteristics in a heterogeneous group of normogonadotropic oligo/amenorrheic infertile women (WHO class 2) for CC-induced follicle growth and ovulation, separately from conception.

In the present study, CCR of 63% within six cycles and 73% within nine ovulatory CC-induced cycles have been reached. Two thirds of patients who conceived reached this end point within the first three ovulatory CC treatment cycles. This is in agreement with previous reports in the literature regarding CC (Garcia *et al.*, 1977) (Gorlitsky *et al.*, 1978) (Gysler *et al.*, 1982) and is similar to spontaneous conception chances in normoovulatory women (Tietze, 1968). This is also comparable to conception rates reported for exogenous gonadotropin induction of ovulation (Dor *et al.*, 1980) (Hamilton-Fairley *et al.*, 1991) (van Santbrink *et al.*, 1995) in anovulatory infertile patients. These observations strongly suggest that the overall detrimental effects of CC on cervical mucus production or endometrial receptivity and subsequent implantation are limited with daily CC doses up to 150 mg. Although the CCR seems to be lower in the high dose CC group, this finding was not statistically significant. It should be noted that the sample size of this group is limited, so actual differences cannot be excluded.

Previous studies were unable to identify predictors for conception in CC induction of ovulation in normogonadotropic infertile women (Hammond *et al.*, 1983) (Shoham *et al.*, 1990) (Kousta *et al.*, 1997). In the present study, age and cycle history (amenorrhea or oligomenorrhea) were significantly different comparing patients who conceived *vs.* those who did not during CC-induced ovulatory cycles. Multivariate analyses revealed a final model including age and cycle history. The predictive power of age was the highest. The area under the receiver operating characteristics curve of the final model including these two factors reached 0.68 (data not shown), which is substantially lower than that in the previous model predicting ovulation after CC (0.82) (Imani *et al.*, 1998). For reasons of clarity, the forward stepwise approach was chosen (see **Table 2**). Backward stepwise analysis was also applied, resulting in the

same final model (data not shown). Young patients have a higher probability to conceive during CC-induced ovulatory cycles. The fecundability rate of the patient decreases by approximately 10% per year. This is in agreement with reports that indicate that age is an important factor for the prediction of chances for spontaneous conception in untreated normoovulatory subfertile patients (Eimers *et al.*, 1994) (Imani *et al.*, 1998) (Scott *et al.*, 1995). Similar findings have been reported for the prediction of chances to conceive following exogenous gonadotropin induction of ovulation (Dor *et al.*, 1980) and *in vitro* fertilization treatment (Templeton *et al.*, 1996).

Amenorrheic patients exhibit a 2-fold higher probability to conceive after ovulatory CC cycles as compared to oligomenorrheic patients. Patients with longer bleeding intervals also exhibit higher conception chances. We have been unable to find similar reports in the literature regarding induction of ovulation. The most likely explanation seems to be the following. Patients with amenorrhea have an extremely low probability to conceive without intervention due to anovulation before seeking help by a physician. Presumably, the major subfertility factor in these patients is chronic anovulation, which can be resolved temporarily by the use of CC medication. These patients are more likely to be at low risk for any other subfertility factor, such as tubal or sperm dysfunction. Some oligomenorrheic patients may have spontaneous ovulations (Minakami *et al.*, 1988), and some of these women may never seek medical intervention because of spontaneous pregnancies. Their benefit from ovulation induction is an increased chance for conception due to an increased number of ovulations with a fixed interval in a given period of time. Similar observations were made for pregnancy chances after artificial insemination with donor sperm in relation to the sperm quality of the partner. Donor insemination outcome is significantly better in cases with very poor sperm quality (Emperaire *et al.*, 1982). One should consider that amenorrheic patients also have a higher probability to remain anovulatory after CC, as demonstrated previously (Imani *et al.*, 1998). In contrast, regular ovulatory cycles are easier to induce with CC in oligomenorrheic patients but there is less chance of pregnancy.

In the present study, patients with elevated initial serum LH levels have a significantly higher probability to conceive once ovulatory cycles have been achieved by CC. A recent study also indicated higher initial LH levels in patients who conceived after CC medication (Kousta *et al.*, 1997). In contrast, a poor treatment outcome has been observed in patients with high LH levels during the follicular phase of CC-induced cycles (Shoham *et al.*, 1990). It should be realized that we report on initial LH levels before initiation of CC medication, rather than during CC-induced cycles. Indeed, elevated LH levels may normalize only during CC-induced ovulatory cycles (Eden *et al.*, 1989). These observations are in sharp contrast with reports regarding patients with elevated LH who perform poorly during gonadotropin induction of ovulation (Hamilton-Fairley *et al.*, 1991) or *in vitro* fertilization (Howles *et al.*, 1986). We previously showed that initial LH concentrations did not predict

patients who would remain anovulatory after CC medication (Imani *et al.*, 1998). The present observation seems to dispute previous beliefs concerning the detrimental effects of raised LH levels on oocyte maturation and capacity for fertilization. In addition, we did not observe a higher spontaneous abortion rate in patients with high LH levels who conceived after CC medication. This finding is also in contrast with previous reports on the effects of high LH concentrations on chances for spontaneous abortion after exogenous gonadotropins (Homburg *et al.*, 1988) (Regan *et al.*, 1990). The predictive power of the initial LH level is poor in case age enters in the final model, which may be due to a correlation of initial LH with age (data not shown). The mechanism of action of CC is not fully elucidated, and the role of LH in the pathogenesis of ovarian abnormalities remains open for speculation. As an example, female siblings of male patients described with an activating mutation of the LH receptor (so-called familial male testotoxicosis) seem to be without a clear phenotype (Fauser *et al.*, 1999)

Total motile sperm count is not a predictor in univariate or multivariate analyses of prediction of conception in the present study population. This is in agreement with the observation that a large overlap exists between semen characteristics of males from fertile *vs.* subfertile couples (Ombelet *et al.*, 1997). Moreover, it can be speculated that couples with better sperm parameters have already conceived spontaneously before seeking medical intervention.

In summary, it can be concluded that body weight and hyperandrogenemia are the predominant predictors for ovulation after CC treatment, whereas age and cycle history dictate pregnancy chances in ovulatory women. This stresses for the first time the important concept that follicle growth and oocyte quality (and subsequent capacity to be fertilized *in vivo*) are differentially regulated during induction of ovulation, confirming observations during *in vitro* fertilization.

References

Adashi EY. Ovulation induction: clomiphene citrate. In: Adashi EY, Rock JA, Rosenwaks Z, editors. *Reproductive endocrinology, surgery, and technology*. Philadelphia: Lippencott-Raven, 1996:1182-1206.

Cox DR. (1972) Regression models and life tables. *J R Statist Soc B.* 34, 187-220.

de Jong FH, Baird DT, van der Molen HJ. (1974) Ovarian secretion rates of oestrogens, androgens and progesterone in normal women and in women with persistent ovarian follicles. *Acta Endocrinol (Copenh).* 77, 575-87.

Dor J, Itzkowic DJ, Mashiach S, Lunenfeld B, Serr DM. (1980) Cumulative conception rates following gonadotropin therapy. *Am J Obstet Gynecol.* 136, 102-5.

Eden JA, Place J, Carter GD, Alaghband-Zadeh J, Pawson ME. (1989) The role of chronic anovulation in the polycystic ovary syndrome: normalization of sex-hormone-binding globulin levels after clomiphene-induced ovulation. *Clin Endocrinol (Oxf).* 30, 323-32.

Eimers JM, te Velde ER, Gerritse R, Vogelzang ET, Looman CW, Habbema JD. (1994) The prediction of the chance to conceive in subfertile couples. *Fertil Steril.* 61, 44-52.

Emperaire JC, Gauzere-Soumireu E, Audebert AJ. (1982) Female fertility and donor insemination. *Fertil Steril.* 37, 90-3.

Fauser BC, Pache TD, Lamberts SW, Hop WC, de Jong FH, Dahl KD. (1991) Serum bioactive and immunoreactive luteinizing hormone and follicle- stimulating hormone levels in women with cycle abnormalities, with or without polycystic ovarian disease. *J Clin Endocrinol Metab.* 73, 811-7.

Fauser BC, Conway GS, Franks S, Genetics of female reproductive dysfunction. Fauser BC, Rutherford, A. J., Strauss, J. F. 3d, and Steirteghem, A. C. Molecular Biology in Reproductive Medicine. Parthenon Publishing. London. Genetics of female reproductive dysfunction. In: Fauser BC, Rutherford AJ, Strauss JFd, Steirteghem AC, editors. *Molecular Biology in Reproductive Medicine.* London: Parthenon Publishing, 1999.

Garcia J, Jones GS, Wentz AC. (1977) The use of clomiphene citrate. *Fertil Steril.* 28, 707-17.

Gorlitsky GA, Kase NG, Speroff L. (1978) Ovulation and pregnancy rates with clomiphene citrate. *Obstet Gynecol.* 51, 265-9.

Gysler M, March CM, Mishell DR, Jr., Bailey EJ. (1982) A decade's experience with an individualized clomiphene treatment regimen including its effect on the postcoital test. *Fertil Steril.* 37, 161-7.

Hamilton-Fairley D, Kiddy D, Watson H, Sagle M, Franks S. (1991) Low-dose gonadotrophin therapy for induction of ovulation in 100 women with polycystic ovary syndrome. *Hum Reprod.* 6, 1095-9.

Hammond MG, Halme JK, Talbert LM. (1983) Factors affecting the pregnancy rate in clomiphene citrate induction of ovulation. *Obstet Gynecol.* 62, 196-202.

Homburg R, Armar NA, Eshel A, Adams J, Jacobs HS. (1988) Influence of serum luteinising hormone concentrations on ovulation, conception, and early pregnancy loss in polycystic ovary syndrome. *Bmj.* 297, 1024-6.

Howles CM, Macnamee MC, Edwards RG, Goswamy R, Steptoe PC. (1986) Effect of high tonic levels of luteinising hormone on outcome of in-vitro fertilisation. *Lancet.* 2, 521-2.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (1998) Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 83, 2361-5.

Kaplan EL, Meier P. (1958) Nonparametric estimation from incomplete observations. *J Am Statist Assoc.* 53, 457-81.

Kousta E, White DM, Franks S. (1997) Modern use of clomiphene citrate in induction of ovulation. *Hum Reprod Update.* 3, 359-65.

Lobo RA, Gysler M, March CM, Goebelsmann U, Mishell Jr DR. (1982) Clinical and laboratory predictors of clomiphene response. *Fertil Steril.* 37, 168-174.

Macgregor AH, Johnson JE, Bunde CA. (1968) Further clinical experience with clomiphene citrate. *Fertil Steril.* 19, 616-22.

Minakami H, Abe N, Izumi A, Tamada T. (1988) Serum luteinizing hormone profile during the menstrual cycle in polycystic ovarian syndrome. *Fertil Steril.* 50, 990-2.

Ombelet W, Bosmans E, Janssen M, Cox A, Vlasselaer J, Gyselaers W, et al. (1997) Semen parameters in a fertile versus subfertile population: a need for change in the interpretation of semen testing. *Hum Reprod.* 12, 987-93.

Opsahl MS, Robins ED, O'Connor DM, Scott RT, Fritz MA. (1996) Characteristics of gonadotropin response, follicular development, and endometrial growth and maturation across consecutive cycles of clomiphene citrate treatment. *Fertil Steril.* 66, 533-9.

Organization WH. (1992) *WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction*, Cambridge University Press,Cambridge.

Pache TD, Wladimiroff JW, Hop WC, Fauser BC. (1992) How to discriminate between normal and polycystic ovaries: transvaginal US study. *Radiology.* 183, 421-3.

Polson DW, Kiddy DS, Mason HD, Franks S. (1989) Induction of ovulation with clomiphene citrate in women with polycystic ovary syndrome: the difference between responders and nonresponders. *Fertil Steril.* 51, 30-4.

Regan L, Owen EJ, Jacobs HS. (1990) Hypersecretion of luteinising hormone, infertility, and miscarriage. *Lancet.* 336, 1141-4.

Rowe PJ, Comhaire FH, Hargreave TB, Mellows HJ. (1993) *WHO manual for the standardized investigation and diagnosis of the infertile couple.*, Cambridge University Press,Cambridge, UK.

Multiple imputation in sample surveys - a phenomenological Baysian approach to nonresponse. Proceedings of the survey Research methods section of the American Statistical Association; 1978.

Schipper I, de Jong FH, Fauser BC. (1998) Lack of correlation between maximum early follicular phase serum follicle stimulating hormone concentrations and menstrual cycle characteristics in women under the age of 35 years. *Hum Reprod.* 13, 1442-8.

Scott RT, Opsahl MS, Leonardi MR, Neall GS, Illions EH, Navot D. (1995) Life table analysis of pregnancy rates in a general infertility population relative to ovarian reserve and patient age. *Hum Reprod.* 10, 1706-10.

Shepard MK, Balmaceda JP, Leija CG. (1979) Relationship of weight to successful induction of ovulation with clomiphene citrate. *Fertil Steril.* 32, 641-5.

Shoham Z, Borenstein R, Lunenfeld B, Pariente C. (1990) Hormonal profiles following clomiphene citrate therapy in conception and nonconception cycles. *Clin Endocrinol (Oxf).* 33, 271-8.

Templeton A, Morris JK, Parslow W. (1996) Factors that affect outcome of in-vitro fertilisation treatment. *Lancet.* 348, 1402-6.

Tietze C. (1968) Fertility after discontinuation of intrauterine and oral contraception. *International journal of fertility.* 13, 385-389.

van Santbrink EJ, Donderwinkel PF, van Dessel TJ, Fauser BC. (1995) Gonadotrophin induction of ovulation using a step-down dose regimen: single-centre clinical experience in 82 patients. *Hum Reprod.* 10, 1048-53.

van Santbrink EJ, Hop WC, Fauser BC. (1997) Classification of normogonadotropic infertility: polycystic ovaries diagnosed by ultrasound versus endocrine characteristics of polycystic ovary syndrome. *Fertil Steril.* 67, 452-8.

9

A nomogram to predict live birth probability after clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility

ABSTRACT

Objective: To establish whether initial screening characteristics of normogonadotropic anovulatory infertile women can aid in predicting live birth after induction of ovulation with clomiphene citrate (CC).
Design: Prospective longitudinal single-center study.
Setting: Specialist academic fertility unit.
Patient(s): Two hundred fifty-nine couples with a history of infertility, oligoamenorrhea, and normal follicle-stimulating hormone (FSH) concentrations who have not been previously treated with any ovulation-induction medication.
Intervention(s): 50, 100, or 150 mg of oral CC per day, for 5 subsequent days per cycle.
Main Outcome Measure(s): Conception leading to live birth after CC administration.
Result(s): After receiving CC, 98 (38%) women conceived, leading to live birth. The cumulative live birth rate within 12 months was 42% for the total study population and 56% for the ovulatory women who had received CC. Factors predicting the chances for live birth included free androgen index (testosterone/sex hormone–binding globulin ratio), body mass index, cycle history (oligomenorrhea versus amenorrhea), and the woman's age.
Conclusion(s): It is possible to predict the individual chances of live birth after CC administration using two distinct prediction models combined in a nomogram. Applying this nomogram in the clinic may be a step forward in optimizing the decision-making process in the treatment of normogonadotropic anovulatory infertility. Alternative first line of treatment options could be considered for some women who have limited chances for success.

INTRODUCTION

Anovulation represents the most frequent cause of female infertility, and in most women normal serum follicle-stimulating hormone concentrations are found. Since its introduction (Greenblatt *et al.*, 1961) clomiphene citrate (CC) has been used worldwide as first choice medication in the treatment of anovulatory infertility. Since the early seventies CC treatment was restricted to normogonadotropic oligoamenorrheic infertility (WHO group 2) (Rowe *et al.*, 1993). A significant proportion of these women, however, remains anovulatory after CC medication. Out of the 75% ovulatory CC patients, approximately 50% will conceive within 6 CC-induced cycles (Imani *et al.*, 1998) (Imani *et al.*, 1999).

Assessment of pregnancy chances after CC induction of ovulation has been the subject of several investigations, all of which focused on the entire group of anovulatory patients who start with CC therapy (Macgregor *et al.*, 1968) (Gorlitsky *et al.*, 1978) (Shepard *et al.*, 1979) (Hammond *et al.*, 1983) (Polson *et al.*, 1989) (Opsahl *et al.*, 1996) (Kousta *et al.*, 1997). These investigators have failed to identify predictors of CC treatment outcome. In contrast, our group focused on prediction of ovulation and conception separately (Imani *et al.*, 1998) (Imani *et al.*, 1999). This approach in life table analysis for prediction of pregnancy chances seems mandatory for the following statistical reasons. Discontinuation of CC therapy due to persistent anovulation is clearly an informative selective drop-out since chances for CC treatment outcome are different compared to women who continue CC therapy. Therefore, inclusion of both CC responders and CC non-responders in a life table analysis for prediction of conception should be avoided.

We developed two distinct prediction models applying multivariate analyses (Imani *et al.*, 1998) (Imani *et al.*, 1999). The first model predicts ovarian response after CC in the entire group of anovulatory patients on the basis of initial screening characteristics such as; free androgen index (testosterone/sex hormone-binding globulin ratio), body mass index (BMI), cycle history (oligomenorrhea versus amenorrhea) and mean ovarian volume (Imani *et al.*, 1998) (Imani *et al.*, 2000) . The second model predicts the chances of conception exclusively in those women who have reached ovulatory cycles after CC and includes woman's age and cycle history (Imani *et al.*, 1999). Although scientifically sound, this approach seems difficult to apply in the daily clinical practice. Moreover, the question remains unanswered whether the chances of having a live birth after CC can be predicted prior to the initiation of medication. By combining both prediction models a nomogram may be developed predicting chances of a given anovulatory patient to conceive resulting in live birth after CC.

Anovulatory patients present with a wide range of chances to conceive after CC. This may be due to differences in the underlying ovarian abnormalities, patient's age, body weight, and individual differences in the anti-estrogenic effects of CC on cervical mucus or endometrium. Applying a nomogram in the clinic may render the ovulation induction protocols more patient tailored and more cost-effective. Patients with a poor

predicted chance to conceive could be advised to refrain from CC therapy and start with an alternative first line treatment modality such as weight reduction, insulin sensitizing agents, or exogenous gonadotropins. Particularly in women of advanced age precious time to ascertain that CC treatment is ineffective may be used for a more effective approach as the first line therapy. We now report the construction of a nomogram to identify characteristics upon initial screening of a large cohort of normogonadotropic anovulatory infertile women predicting the individual chance of pregnancy leading to live birth after CC induction of ovulation.

MATERIALS AND METHODS

Subjects and study protocol

Between February 1993 and May 1999, two hundred and fifty nine patients attending our infertility unit were included in the present study using the following inclusion criteria: (a) oligomenorrhea (bleeding intervals between 35 days and 6 months) or amenorrhea (bleeding interval > 6 months), (b) serum follicle-stimulating hormone (FSH) levels within normal limits (1-10 IU/L) (Rowe *et al.*, 1993) (van Santbrink *et al.*, 1995), (c) normal serum prolactin and thyroid-stimulating hormone levels, (d) spontaneous menses or positive bleeding response to progestagen withdrawal, (e) body mass index (BMI) (weight divided by the square of the patients height) > 18 kg/m^2, (f) between 19-40 years of age, (g) no previous use of ovulation induction agents, (h) a total motile sperm count [TMC = ejaculate volume (milliliters) × sperm concentration (10^6 / mL) × percentage of progressive motile sperm] of the partner above 1 million, (i) negative history of any tubal pathology, and finally (j) no indication for intrauterine insemination. Institutional Review Board approval was obtained from the human subjects committee of the Erasmus University Medical Center Rotterdam and informed consent was obtained from all subjects.

Standardized initial clinical, sonographic, and endocrine screening took place prior to initiation of CC ovulation induction, as described previously (Imani *et al.*, 1998) (Imani *et al.*, 1999) (Imani *et al.*, 2000). Clinical screening included age, type of infertility, cycle history, BMI, waist-to-hip ratio (WHR), and previous medication and/or surgery. Transvaginal sonography (TVS) included assessment of the ovarian stroma echogenicity (arbitrarily classified from 1 to 3 per ovary), ovarian volume (mL) and total number of follicles (both ovaries), as described previously (Pache *et al.*, 1992) (van Santbrink *et al.*, 1997). Sonographic monitoring was performed by a single observer (B.I.). Endocrine screening included serum assays for FSH, luteinising hormone (LH), estradiol (E_2), androstenedione (AD), testosterone (T), and sex-hormone binding globulin (SHBG), fasting insulin and glucose, free and total insulin-like growth factor-I (IGF-I), inhibin B and leptin concentrations, as described previously (Imani *et al.*, 2000). Hormone assays used and the intra and inter assay

coefficients of variation valid for this study have all been described previously (Imani *et al.*, 1998) (Imani *et al.*, 1999) (Imani *et al.*, 2000).

The treatment protocol, assessment of ovarian response and conception after CC have also been described previously (Imani *et al.*, 1998) (Imani *et al.*, 1999). In brief, initial CC doses were 50 mg/day starting on cycle day 3 after a spontaneous or progestagen-induced withdrawal bleeding. In case of absent ovarian response, doses were increased to 100 and 150 mg / day in subsequent cycles. Ovulation after CC medication was assessed by sonographic monitoring of follicle growth and midluteal progesterone > 25 nmol/L. Conception was defined as a positive urinary pregnancy test (clearview, hCG II, Unipath Ltd, Bedford, UK) more than 3 days after the expected menses. Live birth was defined as delivery of a baby. Information regarding deliveries and the health condition of the babies born was collected using the hospital records. In the case of home delivery we collected information directly from the patient and her general practitioner or midwife.

Data analysis

The statistical analysis of conception leading to live birth in anovulatory patients after CC should take into account the following two steps: ovulatory response to CC and conception in case of ovulation. Patients remaining anovulatory after CC (CRA = clomiphene resistant anovulation) are considered to have no chance to conceive with this therapy. Therefore, the cumulative rate of conception leading to live birth was calculated by multiplying the chance of achieving ovulation after CC, with the estimated Kaplan-Meier (Kaplan and Meier, 1958) cumulative probabilities for conception in the group of ovulatory women after CC.

A prediction model of the probability of conception resulting in live birth within 6 months after initiation of CC ovulation induction was constructed by combining the previously published prediction model for the chance to be CRA (Imani *et al.*, 1998) with a prediction model for the probability of conception leading to live birth in ovulatory patients after CC. The current study differs from the previously published model predicting chances for conception (Imani *et al.*, 1999) with respect to the starting point of follow-up (initiation of CC therapy versus first ovulation after CC). In order to implement these findings in clinical practice we included months rather than ovulatory cycles and live birth instead of ongoing pregnancy in the current analysis. A nomogram was constructed by combining the model to predict ovulation in the entire group of women with the model predicting live birth in ovulatory CC patients.

To determine the goodness-of-fit of the combined prediction model, patients were divided into 5 equal groups according to quintiles of the predicted probability of conception leading to live birth. Using a Chi-square test, the mean predicted probability within each group was compared to the observed probability, calculated by the Kaplan-Meier method described in the previous paragraph. Data were analyzed using the commercially available software package SPSS (Chicago, Ill, USA).

RESULTS

From the total number of 259 patients fulfilling the in/exclusion criteria and treated with CC, 186 (72%) suffered from primary infertility, and 62 (24%) from amenorrhea. One hundred and ten patients (43%) were obese (BMI > 27), 160 patients (62%) presented with an elevated FAI (FAI > 4.5) and 100 patients (39%) with hyperandrogenemia (T ≥ 3.2 nmol/L and/or AD ≥ 16.3 nmol/L (Fauser *et al.*, 1991) (van Santbrink *et al.*, 1997)). In 195 patients (75%) polycystic ovaries (mean ovarian volume ≥ 10.8 mL and/or mean follicle number per ovary ≥ 10 (Pache *et al.*, 1992) (van Santbrink *et al.*, 1997)) were diagnosed. One hundred and twenty-one patients (47%) presented with elevated initial LH (LH ≥ 7.0 IU/L (van Santbrink *et al.*, 1997)) serum level. Sixty-five patients (25%) remained anovulatory after CC medication despite the maximum CC dose (150 mg/day) and were considered as CRA. Eighty-three patients (32%) presented with ovulatory cycles after CC but failed to conceive.

A total number of 111 (43%) women conceived after CC treatment and 11 (10% of conceptions) miscarried. There was one case of ectopic pregnancy and one case of intrauterine death. A total number of 98 (38%) patients had a live birth. Sixty-seven patients delivered at the hospital and could leave the unit within 24 hours. Nineteen patients were hospitalized for pre and postnatal care. In 4 pregnant women anti-hypertensive medication was required. There was no case of maternal or fetal death. In 15 women an instrumental delivery and in 13 patients a cesarean section was performed. At birth the mean ± SD gestational age was 39.3 ± 2.8 weeks and the mean weight of the babies was 3142 ± 590 gram. There was no case of postpartum death in

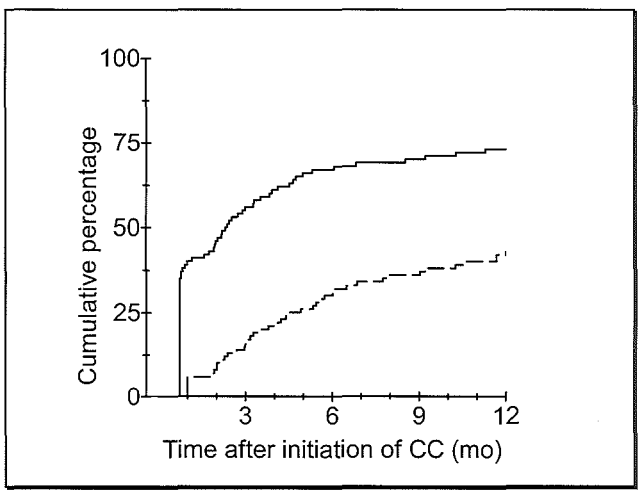


Figure 1 Cumulative percentage of women reaching first ovulation and first conception ending in a live birth in 259 normogonadotropic oligoamenorrheic infertile patients treated with CC. Solid line = first ovulation after CC (n=194; 75%); dashed line = first conception ending in a live birth (n=98; 38%).

the first year after the delivery. No case of congenital abnormalities was reported.

Odds ratios of screening characteristics separately for ovulation in the entire population and live birth in ovulatory CC women are presented in **Table 1**. Screening characteristics were included on the basis of significance in previous multivariate analysis predicting ovulation (FAI, BMI, cycle history, ovarian volume, leptin), predicting pregnancy in ovulatory patients (age, cycle history) and biological relevance for ovarian dysfunction (LH, hyperandrogenemia, insulin, insulin/glucose ratio, inhibin B, free IGF-I) or pregnancy prediction (infertility history, bleeding interval in case of oligomenorrhea, total motile sperm count). The cumulative rates of

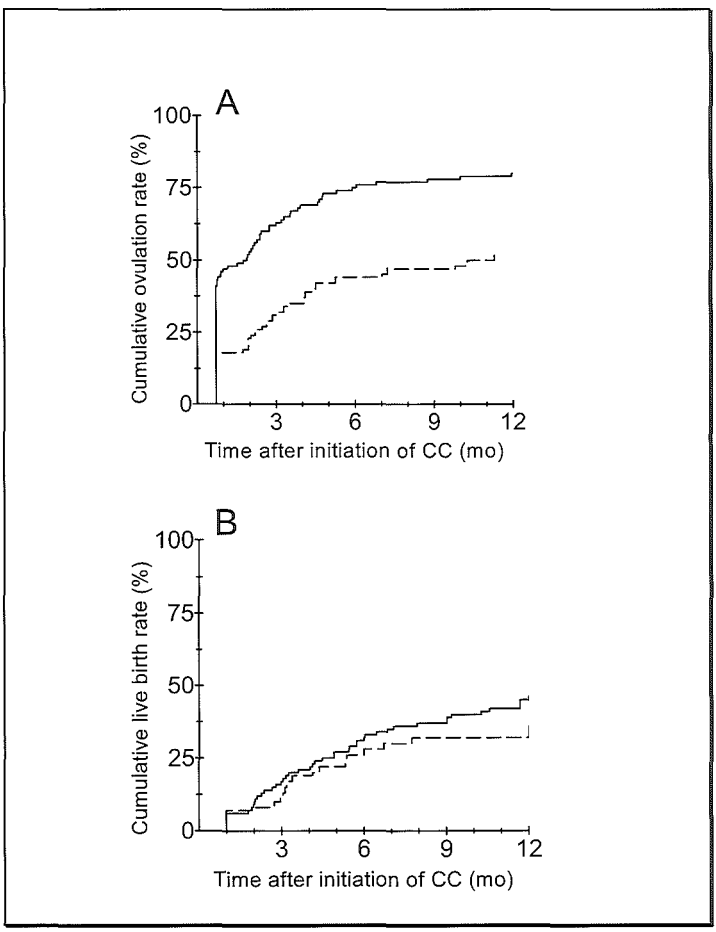


Figure 2: The impact of cycle history on cumulative ovulation and live birth rates in 259 normogonadotropic oligoamenorrheic infertile patients treated with CC. A: solid line = oligomenorrheic patients (n = 197); B: dashed line = amenorrheic patients (n = 62). P value represents differences applying LogRank test (A: $P < 0.001$; B: P value = not applicable in the overall WHO 2 anovulatory patients. $P < 0.001$ in ovulatory patients following CC).

first ovulation and first conception ending in live birth are depicted in **Figure 1**. The impact of cycle history on cumulative ovulation and live birth rates after CC is depicted in **Figure 2**.

Prediction of conception leading to live birth after CC applying a nomogram has been depicted in **Figure 3**. This nomogram comprises two separate steps. In the left panel the probability of ovulation after CC can be estimated applying FAI, BMI, and cycle history (oligomenorrhea versus amenorrhea). Drag and drop the chance to ovulate after CC from the left to the right panel. In the right panel the predicted probability to conceive within 6 months after initiation of CC medication resulting in live birth is indicated, combining the probability of ovulation after CC with age and cycle history (oligomenorrhea versus amenorrhea) of the patient. Mean ovarian volume, and serum leptin levels comprise the least predictive contribution to the model (Imani *et al.*, 1998) (Imani *et al.*, 2000) and therefore were excluded from the left panel of this nomogram.

The P-values for the Chi-square goodness-of-fit test was 0.49, indicating no statistically significant lack of fit between the observed and predicted probability of live birth within the 5 groups (see also Materials & Methods; data not shown).

DISCUSSION

Currently, CC represents the first line treatment strategy for all patients with normogonadotropic anovulatory infertility. The present study confirms that CC is safe and convenient with limited chances for complications such as multiple pregnancy or OHSS. However, only 38% of treated women may conceive after CC resulting in live birth which is substantially less than generally assumed. It may represent a distinct step forward if patients with even lower chances for live birth after CC could be identified in advance.

Previous attempts by retrospective studies in the entire group of anovulatory patients failed to identify factors predicting ovulation (Gorlitsky *et al.*, 1978) (Shepard *et al.*, 1979) (Polson *et al.*, 1989) or conception (Gorlitsky *et al.*, 1978) (Hammond *et al.*, 1983) (Kousta *et al.*, 1997) after CC. In contrast, we focused on predictors of CC treatment outcome in a prospective fashion in the current longitudinal cohort study. For the first time we used ovulation and conception as separate endpoints, taking into account that a significant proportion of patients remains anovulatory during CC medication. These women are not exposed to the chances of becoming pregnant. Univariate and multivariate analyses were performed in these two separate settings. The first multivariate model consisting of FAI, BMI, cycle history, and mean ovarian volume predicts chances of remaining anovulatory after CC (Imani *et al.*, 1998). The second model consisting of age and cycle history predicts chances for conception only in women who respond to CC (Imani *et al.*, 1999). Observed screening characteristics

Table 1: Odds ratios (OR) of initial screening characteristics in relation to the occurrence of ovulation in the entire group of normogonadotropic oligoamenorrheic infertile women (n=194 of 259) or live birth in ovulatory women after CC (n=98 of 194)

Screening characteristics	Ovulation following CC	Live birth in ovulatory women following CC
Clinical		
Age (per yr) [a]	1.05 (0.98 – 1.12)	0.90 (0.85 – 0.94)
Primary vs. secondary infertility [b]	1.23 (0.64 – 2.34)	0.71 (0.44 – 1.15)
Oligomenorrhea vs. amenorrhea [a, c, d]	4.34 (2.34 – 8.05)	0.77 (0.47 – 1.28)
Bleeding interval in case of oligomenorrhea [b, e]		
5-6 weeks	8.30 (3.29 –20.98)	0.65 (0.36 – 1.18)
6-9 weeks	3.64 (1.75 – 7.55)	0.87 (0.50 – 1.51)
9-26 weeks	2.90 (1.25 – 6.73)	0.83 (0.42 – 1.63)
BMI (kg/m2) [c]	0.92 (0.88 – 0.96)	1.00 (0.97 – 1.04)
Waist to hip ratio (WHR per 0.1) [b]	0.60 (0.40 – 0.89)	1.08 (0.81 – 1.42)
Endocrine		
LH (IU/L) [b]	0.97 (0.91 – 1.03)	1.04 (1.00 – 1.09)
FAI [b, f]	0.83 (0.78 – 0.89)	1.06 (1.01 – 1.11)
Hyperandrogenemia (elevated T and / or AD) [b]	0.36 (0.20 – 0.65)	1.67 (1.10 – 2.54)
Insulin (mU/L) [b]	0.95 (0.92 – 0.98)	0.99 (0.97 – 1.02)
Insulin/glucose ratio [b]	0.87 (0.77 – 0.97)	0.99 (0.90 – 1.08)
Inhibin B (ng/L) [b]	1.00 (0.99 – 1.01)	1.00 (0.99 – 1.01)
Free IGF-I (ng/mL) [b]	0.91 (0.77 – 1.09)	1.10 (0.96 – 1.25)
Leptin (ng/mL) [c]	0.97 (0.95 – 0.99)	1.02 (1.00 – 1.03)
Ultrasound		
PCO versus non-PCO [c, g]	0.59 (0.29 – 1.22)	0.98 (0.63 – 1.54)
Total motile sperm count / 10^4 [b]	-	1.00 (0.81 – 1.23)

[a] Characteristics predicting conception in ovulatory CC patients applying multivariate analysis as published previously (Imani et al., 1999)

[b] Biological relevant characteristics for ovarian dysfunction or pregnancy prediction in normogonadotropic anovulation

[c] Characteristics predicting ovulation following CC applying multivariate analysis as published previously (Imani et al., 1998) (Imani et al., 2000)

[d] The oligomenorrheic patients are 4 times more likely (OR = 4.34) to ovulate after CC compared to amenorrheic patients. However the chance for a live birth after CC is 23% less (OR = 0.77) compared to amenorrheic patients.

[e] Amenorrhea is reference category for the categories of bleeding interval in oligomenorrhea.

[f] FAI = T × 100 / SHBG

[g] Polycystic ovaries = mean ovarian volume ≥ 10.8 mL and/or mean follicle number per ovary ≥ 10 (van Santbrink et al., 1997)

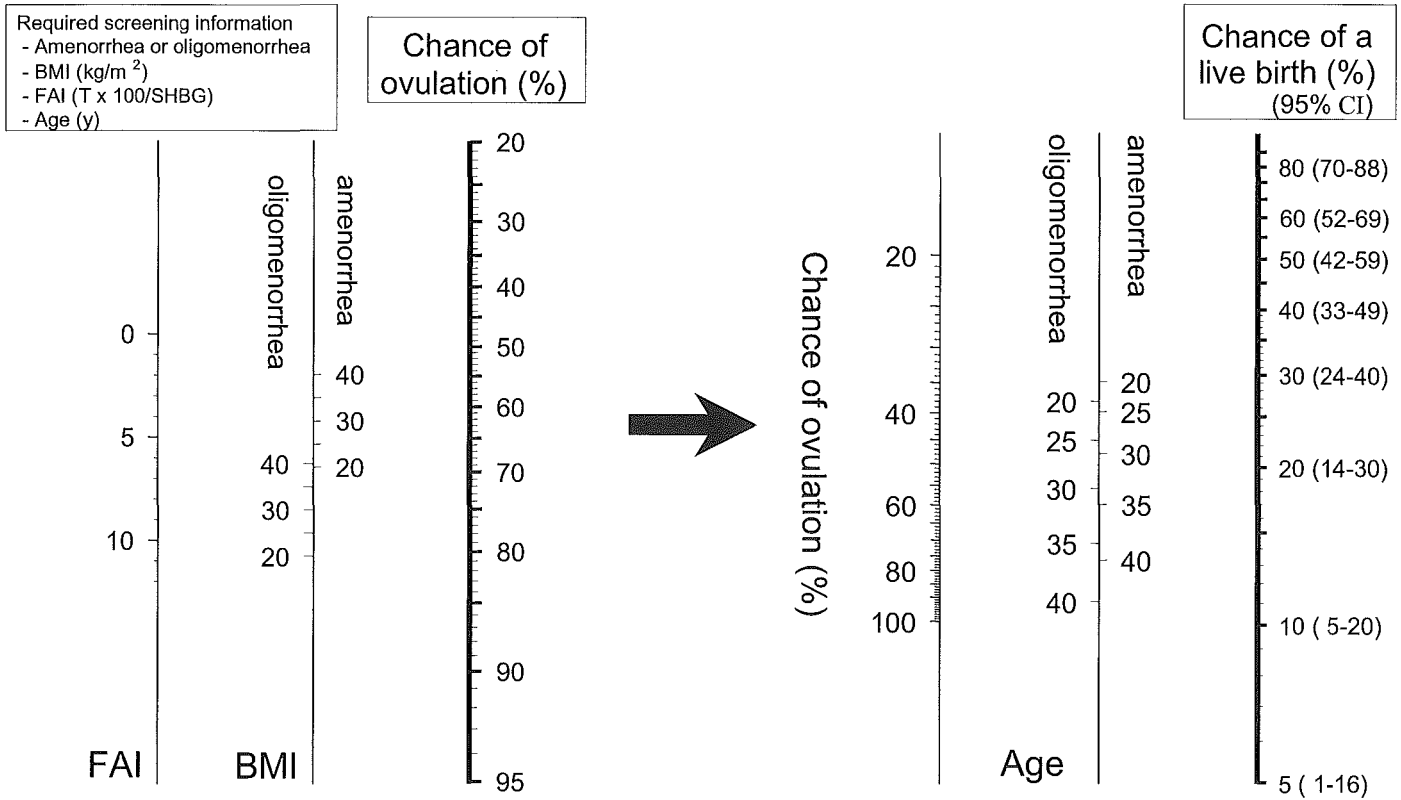


Figure 3 The probability to ovulate and conceive ending in a live birth within 6 months after initiation of CC therapy applying a nomogram. Chances for ovulation following CC can be assessed using FAI, BMI, and cycle history (*left panel*). The probability to conceive ending in a live birth within 6 months after initiation of CC can be assessed by combining the probability to ovulate with the age and the cycle history of the woman (*right panel*).Encircle the values related to each screening parameter. Connect the circles to observe the predicted probability to ovulate. Correspond this probability from the left to the right panel and connect it with the age of the woman (separately for oligomenorrhea or amenorrhea) to predict the chance of a live birth. For instance, a 29 year old amenorrheic patient presented with the following findings: FAI = 9.3, BMI = 32. Her chance to conceive leading to a live birth within 6 months after initiation of CC medication will be 19% (chance to ovulate following CC: 40%).

involved in the prediction of ovulation after CC medication are distinctly different from predictors of conception in ovulatory CC patients. In the current study we combine both prediction models in a nomogram to predict chances of live birth in the entire group of normogonadotropic anovulatory patients. For reasons of its applicability in clinical practice we converted analysis to month rather than treatment cycles and used live birth rather than pregnancy as the outcome. The estimation of chances to ovulate is a crucial step in evaluating the probability of conception resulting in live birth after CC therapy. This is the reason for the involvement of two steps in the presented nomogram. Cycle history is the only factor which is present in both steps. Amenorrheic patients are less likely to ovulate after CC medication, but once they ovulate chances for live birth are higher compared to oligomenorrheic women. A biologically plausible explanation is that the occurrence of other factors contributing to a reduced fertility is less likely since these women never had the chance to conceive. BMI does not influence live birth rate in patients who ovulate following CC. Moreover, live birth rate in CC-ovulatory PCO *vs.* CC-ovulatory non-PCO patients are comparable. Interestingly, age does not affect chances for ovulation, but is involved in pregnancy chances once ovulation has been reached (Imani *et al.*, 1998) (Imani *et al.*, 1999) . This is in agreement with reports concerning an association between reduced treatment outcome after *in vitro* fertilization and patient's advanced age (Scott *et al.*, 1995)

The major aim was to render the present nomogram as easy as possible for use in daily clinical practice. Therefore, ovarian volume (which was included in the previous multivariate model predicting ovulation (Imani *et al.*, 1998) has been excluded from the present nomogram. The additional predictive contribution of this prognostic factor to the model is negligible (a decrease in the area under the receiver characteristic curve of the model to predict patients remaining anovulatory from 0.82 to 0.80 (Imani *et al.*, 1998). We also demonstrated previously that serum leptin concentrations (an endocrine marker associated with obesity (Rouru *et al.*, 1997)) predict chances of remaining anovulatory after CC (Imani *et al.*, 2000). Leptin has also been excluded from the present nomogram since this assay is not available in most hospital laboratories. Instead, information for BMI can be generated easily with only a minor decrease in AUC of the prediction model (0.85 instead of 0.82). Therefore, serum T and SHBG levels are the only endocrine factors that should be assessed for the present nomogram. Indeed, both obesity and hyperandrogenemia are associated with polycystic ovary syndrome known to exhibit low chances for success during CC ovulation induction. Remaining biological plausible factors involved in ovarian dysfunction in normogonadotropic anovulation such as LH (Fauser *et al.*, 1991), insulin/glucose ratio's (Poretsky *et al.*, 1999), inhibin B (Anderson *et al.*, 1998) and free IGF-I (Thierry van Dessel *et al.*, 1999) failed to predict ovulation or conception and were therefore excluded from the nomogram.

In summary, the present study demonstrates for the first time that a nomogram can be developed on the basis of initial screening characteristics predicting chances of live birth after CC for a given patient. An external validation of the present nomogram is mandatory to be able to define a clear cut-off level in the chances for live birth after CC for the decision making in routine daily clinical practice. For instance, a cut-off level of 20% chance of having a live birth (which comprises 19% of the overall WHO 2 anovulatory patients) could be chosen. These patients could be advised to refrain from CC therapy and start with an alternative first line treatment modality such as weight reduction, exogenous gonadotropins, insulin sensitizing hormones, or *in vitro* fertilization particularly in women with an advanced age. This would render CC ovulation induction strategies more patient tailored and could improve overall cost effectiveness.

References

Anderson RA, Groome NP, Baird DT. (1998) Inhibin A and inhibin B in women with polycystic ovarian syndrome during treatment with FSH to induce mono-ovulation. *Clin Endocrinol (Oxf).* 48, 577-84.

Fauser BC, Pache TD, Lamberts SW, Hop WC, de Jong FH, Dahl KD. (1991) Serum bioactive and immunoreactive luteinizing hormone and follicle- stimulating hormone levels in women with cycle abnormalities, with or without polycystic ovarian disease. *J Clin Endocrinol Metab.* 73, 811-7.

Gorlitsky GA, Kase NG, Speroff L. (1978) Ovulation and pregnancy rates with clomiphene citrate. *Obstet Gynecol.* 51, 265-9.

Greenblatt RB, Barfield WE, Jungck EC, Ray AW. (1961) Induction of ovulation with MRL/41: preliminary report. *J Am Med Assoc.* 178, 101-4.

Hammond MG, Halme JK, Talbert LM. (1983) Factors affecting the pregnancy rate in clomiphene citrate induction of ovulation. *Obstet Gynecol.* 62, 196-202.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (1998) Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 83, 2361-5.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (1999) Predictors of chances to conceive in ovulatory patients during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 84, 1617-22.

Imani B, Eijkemans MJ, de Jong FH, Payne NN, Bouchard P, Giudice LC, et al. (2000) Free androgen index and leptin are the most prominent endocrine predictors of ovarian response during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 85, 676-82.

Kaplan EL, Meier P. (1958) Nonparametric estimation from incomplete observations. *J Am Statist Assoc.* 53, 457-81.

Kousta E, White DM, Franks S. (1997) Modern use of clomiphene citrate in induction of ovulation. *Hum Reprod Update.* 3, 359-65.

Macgregor AH, Johnson JE, Bunde CA. (1968) Further clinical experience with clomiphene citrate. *Fertil Steril.* 19, 616-22.

Opsahl MS, Robins ED, O'Connor DM, Scott RT, Fritz MA. (1996) Characteristics of gonadotropin response, follicular development, and endometrial growth and maturation across consecutive cycles of clomiphene citrate treatment. *Fertil Steril.* 66, 533-9.

Pache TD, Wladimiroff JW, Hop WC, Fauser BC. (1992) How to discriminate between normal and polycystic ovaries: transvaginal US study. *Radiology.* 183, 421-3.

Polson DW, Kiddy DS, Mason HD, Franks S. (1989) Induction of ovulation with clomiphene citrate in women with polycystic ovary syndrome: the difference between responders and nonresponders. *Fertil Steril.* 51, 30-4.

Poretsky L, Cataldo NA, Rosenwaks Z, Giudice LC. (1999) The insulin-related ovarian regulatory system in health and disease. *Endocr Rev.* 20, 535-82.

Rouru J, Anttila L, Koskinen P, Penttila TA, Irjala K, Huupponen R, et al. (1997) Serum leptin concentrations in women with polycystic ovary syndrome. *J Clin Endocrinol Metab.* 82, 1697-700.

Rowe PJ, Comhaire FH, Hargreave TB, Mellows HJ. (1993) *WHO manual for the standardized investigation and diagnosis of the infertile couple.*, Cambridge University Press,Cambridge, UK.

Scott RT, Opsahl MS, Leonardi MR, Neall GS, Illions EH, Navot D. (1995) Life table analysis of pregnancy rates in a general infertility population relative to ovarian reserve and patient age. *Hum Reprod.* 10, 1706-10.

Shepard MK, Balmaceda JP, Leija CG. (1979) Relationship of weight to successful induction of ovulation with clomiphene citrate. *Fertil Steril.* 32, 641-5.

Thierry van Dessel HJ, Lee PD, Faessen G, Fauser BC, Giudice LC. (1999) Elevated serum levels of free insulin-like growth factor I in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 84, 3030-5.

van Santbrink EJ, Hop WC, van Dessel TJ, de Jong FH, Fauser BC. (1995) Decremental follicle-stimulating hormone and dominant follicle development during the normal menstrual cycle. *Fertil Steril.* 64, 37-43.

van Santbrink EJ, Hop WC, Fauser BC. (1997) Classification of normogonadotropic infertility: polycystic ovaries diagnosed by ultrasound versus endocrine characteristics of polycystic ovary syndrome. *Fertil Steril.* 67, 452-8.

10

High Singleton Live Birth Rate Following Classical Ovulation Induction in Normogonadotrophic Anovulatory Infertility (WHO 2)

ABSTRACT

BACKGROUND: Medical induction of ovulation using Clomiphene citrate (CC) as first line and exogenous gonadotrophins as second line forms the classical treatment algorithm in normogonadotrophic anovulatory infertility. Because the chances of success following classical ovulation induction are not well established, a shift in first line therapy can be observed towards alternative treatment. Our aim was to (1) reliably assess the probability of singleton live birth following classical induction of ovulation and (2) construct a prediction model, based on individual patient characteristics assessed upon standardised initial screening, to help identify patients with poor chances of success. METHODS: Two hundred and forty consecutive women visiting a specialist academic fertility unit with a history of infertility, oligomenorrhea or amenorrhea, and normal follicle-stimulating hormone (FSH) and estradiol (E_2) serum concentrations (WHO group 2) were prospectively followed. The women were not treated before with ovulation inducing agents. All patients commenced with CC. Patients who did not ovulate within 3 treatment cycles of incremental daily doses up to 150 mg for 5 consecutive days or ovulatory CC patients who did not conceive within 6 cycles, subsequently underwent gonadotrophin induction of ovulation applying a step-down dose regimen. The main outcome measure was pregnancy resulting in singleton live birth. Cox regression was used to construct a multivariable prediction model. RESULTS: Overall, 134 pregnancies ending in a singleton live birth occurred (56% of women). The cumulative pregnancy rate after 12 and 24 months of follow-up was 50% and 71%, respectively. PCOS patients (49%), clearly non-PCOS patients (13%) and the in-between group did not differ in prognosis (P = 0.9). The multivariable Cox regression model contained the woman's age, the insulin/glucose ratio and the duration of infertility. With a cut-off value of 30% for low chance, the model predicted probabilities at 12 months lower than this cut-off for 25 out of 240 (10%) patients.

CONCLUSIONS: Classical ovulation induction gives very good results in normogonadotrophic anovulatory infertility. Alternative treatment options may not be indicated as first line therapy in these patients, except for subgroups with poor prognosis. These women can be identified by older age, longer duration of infertility and higher insulin/glucose ratio.

INTRODUCTION

Anovulation is a common cause of infertility, diagnosed in at least 25% of couples with fertility problems. In the great majority of these women the underlying cause is described as 'pituitary-ovarian disbalance', where serum follicle-stimulating hormone (FSH) and estradiol (E_2) levels are within normal limits (also described as World Health Organization (WHO) group 2) (World Health Organization, 1993), (The ESHRE Capri Workshop Group, 1995). Patients suffering from the polycystic ovary syndrome (PCOS) are considered a subgroup of WHO 2 (Laven *et al.*, 2002). Since the early 60s patients can be treated effectively with the anti-estrogen Clomiphene Citrate (CC) as well as exogenous gonadotrophins. CC is generally applied as first line treatment in these women, due to low costs and minor chances for side effects or complications. Traditionally, exogenous gonadotrophins (especially FSH) is considered the second line therapy in case of failure to ovulate or conceive following CC. This treatment modality requires frequent monitoring due to inherent risks of multiple follicle development resulting in increased chances for ovarian hyperstimulation syndrome and multiple gestation, especially in PCOS patients (Messinis and Milingos, 1997), (Legro, 1998). Over the years, numerous (mainly retrospective) studies have established the overall effectiveness of either CC (Hammond *et al.*, 1983), (Kousta *et al.*, 1997) or FSH (White *et al.*, 1996), (van Santbrink and Fauser, 1997), (Fauser and Van Heusden, 1997) in varying groups of patients. However, data concerning a prospective follow-up of a coherent CC and FSH treatment strategy in a well-defined group of patients are not yet available.

Over the last 10 years, increased attention is focussed towards alternative treatment options such as the insulin sensitizing agents (Heard *et al.*, 2002), (Nestler *et al.*, 2002), (Glueck *et al.*, 2002), aromatase inhibitors (Mitwally and Casper, 2001), laparoscopic surgery of ovaries (Cohen, 1996), (Farquhar *et al.*, 2001) or assisted reproduction such as intra- uterine insemination (IUI) or in-vitro fertilization (IVF) (Buyalos and Lee, 1996), (Child *et al.*, 2001). Most promising results of up to 80% pregnancies have been reported by individual studies, but the majority of these reports involves a limited number of selected patients, and is retrospective and uncontrolled. It is often suggested that the multiple birth rate is reduced in comparison with ovulation induction, but reliable information on multiple births is mostly lacking.

There seems to be an obvious need to firmly establish the success and complication rates of the conventional approach for ovulation induction (involving CC and FSH) in

a large unselected cohort of patients. The current report contains an integrated analysis of chances for pregnancies resulting in singleton live childbirth, extending previously reported outcomes separately for CC (Imani *et al.*, 2002b) and FSH (Imani *et al.*, 2002a), (Mulders *et al.*, 2003). In addition, a prediction model is presented to help identify patients with poor chances of success. This information may serve as point of reference for future studies involving alternative approaches for ovulation induction.

SUBJECTS AND METHODS

Subjects

Between November 1992 and May 1999, a consecutive series of 240 women attending our infertility unit were included in the present study using the following inclusion criteria: [1] oligomenorrhea (bleeding intervals between 35 days and 6 months) or amenorrhea (bleeding interval > 6 months), [2] normal serum FSH (1-10 IU/l) and E_2 (150-400 pmol/l) levels (van Santbrink *et al.*, 1997), [3] normal serum prolactin and thyroid-stimulating hormone levels, [4] spontaneous menses or positive bleeding response to progestagen withdrawal, [5] body mass index (BMI: weight divided by the square of the patient's height) > 18 kg/m^2, [6] between 19 and 40 years of age, [7] no previous use of ovulation induction agents, [8] a total motile sperm count (TMC = ejaculate volume [ml] × sperm concentration [10^6/ml] × percentage of progressive motile sperm) of the partner above 1 million, [9] negative history of or screening for any tubal pathology, and finally [10] no indication for IUI. Institutional review board approval was obtained from the human subjects committee of the Erasmus Medical Center and informed consent was obtained from all study participants.

The standardized initial clinical, sonographic and endocrine screening took place prior to the initiation of CC ovulation induction, as had been described previously (Imani *et al.*, 2002b). The clinical screening included patient's age, type of infertility (primary or secondary), cycle history (amenorrhea or oligomenorrhea), BMI, waist-to-hip ratio (WHR), and previous medication and/or surgery. Transvaginal sonography included assessment of ovarian volume (ml), and total number of follicles, as described previously (Pache *et al.*, 1992). Endocrine screening included serum assays for FSH, luteinizing hormone (LH), E_2, androstenedione (AD), testosterone (T), sex-hormone binding globulin (SHBG), and fasting insulin and glucose. Hormone assays used and the intra-assay and inter-assay coefficients of variation valid for this study have all been described previously (Imani *et al.*, 2000).

The treatment protocol and the assessment of ovarian response and pregnancy after CC administration have also been described previously (Imani *et al.*, 2002b). In brief, the women received initial CC doses of 50 mg/day from cycle day 3 until day 7 after a spontaneous or progestagen-induced withdrawal bleeding. In cases of absent ovarian response, the daily dosage was increased to 100 and 150 mg in subsequent cycles. Ovulation after CC or FSH medication was assessed by sonographic monitoring of

disappearance of the preovulatory follicle and/or the finding of midluteal progesterone levels > 25 nmol/L. In case of CC resistant anovulation (CRA), i.e. absent ovulation after 3 months (despite stimulation with a maximum daily CC dose of 150 mg), or if pregnancy failed to occur within 6 ovulatory cycles, ovulation induction with exogenous recombinant FSH (Puregon®, Organon NV, Oss, The Netherlands) was applied as second line treatment, using a decremental dose regimen as published previously (van Santbrink and Fauser, 1997). The FSH response dose was assessed during the first cycle applying a low-dose, step-up regimen (Imani *et al.*, 2002a), (Mulders *et al.*, 2003).

Pregnancy was defined as a positive urinary pregnancy test (clearview, hCG II, Unipath Ltd, Bedford, UK) at least 3 days after the expected menses. Ongoing pregnancy was defined as sonographic assessment at 12 weeks amenorrhea of an intrauterine gestational sac with positive heartbeat. Live birth was defined as the delivery of a baby after at least 28 weeks gestational age. Information regarding deliveries and the health condition of the babies born was collected using the hospital records. In cases of home delivery, we collected the information directly from the patient and her general practitioner or midwife.

Statistical analysis

The Kaplan-Meier method was used to estimate the cumulative pregnancy rate leading to singleton live birth. The length of the period from start of CC treatment until the first pregnancy leading to live birth was the time variable in this analysis. Conceptions that ended in miscarriage were ignored by the analysis, and in these cases follow-up continued until pregnancy resulting in singleton live birth occurred. Patients who had a pregnancy ending in premature birth, stillbirth or multiple live birth, were considered censored at conception. Patients who did not become pregnant were censored at their time of last treatment. Spontaneous pregnancies resulting in singleton live birth that occurred during the first or second month after the end of treatment were also included in the analysis.

Cox regression was used for univariable and multivariable analysis relating initial screening characteristics to the cumulative pregnancy rate leading to singleton live birth. In univariable analysis the log-rank test was used, and a P-value < 0.05 was considered as statistically significant. A multivariable prediction model was constructed using a backward stepwise elimination method with $P < 0.15$ for exclusion of predictors, corresponding to Akaike's Information Criterion (Akaike, 1973). Predictor variables with univariable $P < 0.5$ were candidate variables for the multivariable model. These 'liberal' P-values were chosen because the resulting model may be expected to have a better predictive performance in new patients than when the strict $P < 0.05$ criterion is used (Harrell *et al.*, 1996). The prediction model was corrected by a shrinkage factor, determined by a bootstrapping technique with 200 replications, to provide better predictions in new patients (Van Houwelingen and Le Cessie, 1990), (Harrell *et al.*, 1996). The discriminative ability of the prediction model

was assessed by the c-statistic, which is similar to the area under the receiver operating characteristic (ROC) curve for dichotomous outcomes, and can range from 0.5 to 1.0 (Harrell *et al.*, 1985). The bootstrap technique was used again, this time to correct the optimism in the apparent c-statistic (Harrell *et al.*, 1996). Missing values occurred for some screening characteristics. The technique of multiple imputation was used to avoid loss of information due to missing data in multivariable analyses (Little, 1992). Missing data occurred in the WHR (45% missing), semen analysis (25% missing) and in the insulin-glucose ratio (26% missing). Statistical analyses were performed with SPSS for Windows (version 10) and S+ 2000.

RESULTS

Following induction of ovulation, 162 pregnancies occurred in 159 (66%) patients. One hundred and forty seven (90% of pregnancies) were ongoing pregnancies (positive heart action upon ultrasound at 12 weeks amenorrhea), whereas 15 miscarried within 12 weeks. Of the 147 ongoing pregnancies, 137 were singleton and 10 were multiple (7% of ongoing pregnancies). Of the singleton ongoing pregnancies, two ended in premature birth and one in stillbirth. All other ongoing pregnancies ended in live birth (134 singleton and 10 multiple). The cumulative pregnancy rate resulting in singleton live birth is shown in **Figure 1**. After 6, 12, 18 and 24 months of

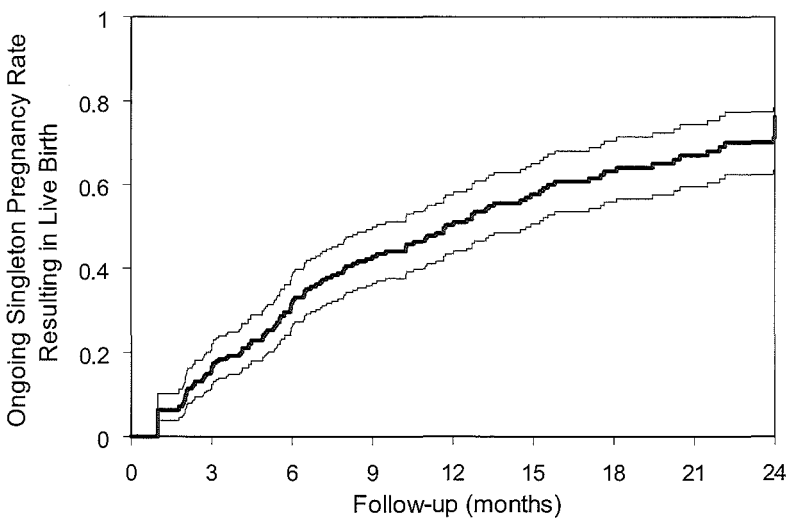


Figure 1: Cumulative pregnancy rate resulting in singleton live birth of a consecutive series of 240 normogonadotrophic anovulatory infertile women undergoing classical ovulation induction (CC as first line, followed by FSH as second line therapy if required), calculated by the Kaplan-Meier method, with 95% Confidence interval.

follow-up, respectively 32%, 50%, 63% and 71% of patients presented with a singleton pregnancy resulting in live birth of a baby. The cumulative chance after 24 months increased to 74% in case multiple live births were included, and to 76% for all ongoing pregnancies. The median time taken to reach pregnancy resulting in singleton live birth was 11.7 months. The median follow-up of the women who did not reach this endpoint was 10 months, with a maximum follow-up of 4 years. Their median duration of CC-treatment was 5.9 months (range 1 to 18) during which a median number of 5.0 (range 1 –13) CC-cycles were performed. The median duration of FSH treatment of the women who did not reach the endpoint was 7.0 months (range 1 to 34) with a median number of 5.0 (range 1 –13) FSH treatment cycles.

Figure 2 shows the fate of the women participating in the study. All 240 women started with CC, and 57 (24%) of them were found to be CRA. CC outcome in the remaining 183 women was 112 pregnancies: 98 ongoing singleton and 4 ongoing twin pregnancies (4% of CC-pregnancies) resulting in live birth, 11 pregnancies ending in miscarriage (10% of CC-pregnancies) and 1 ending in stillbirth. There was one additional miscarriage, in a woman who became pregnant again after a first miscarriage. There were 84 women who started FSH induction of ovulation after unsuccessful CC-treatment. In 44 women an ongoing pregnancy was subsequently achieved, resulting in 36 singleton, 4 twin, one triplet and one quadruplet live birth (14% multiple births) and 2 stillbirths. There were 3 pregnancies that miscarried (6% of 47 pregnancies). The miscarriage rate was not statistically significant between CC and FSH (P = 0.7). Almost 50% (33 out of 71) of the CCF patients did not proceed to FSH treatment and should be considered as dropout. They did not differ from the patients who started FSH treatment in any of the screening characteristics, except for the level of endogenous FSH: the mean level was 4.1 (SD 1.7) IU/l versus 5.1 (SD 1.6) IU/l in the group that started FSH treatment (P = 0.01).

The primary screening characteristics are presented in **Table 1** for women who did reach a pregnancy leading to singleton live birth and for those who did not. Successful patients were younger, infertile for a shorter period of time, thinner, had marginally higher FSH, higher LH and lower insulin/glucose ratios. The final column indicates the screening characteristics that had an independent contribution to the multivariable prediction model: age, the insulin/glucose ratio and the duration of infertility. Figure 3 shows the distribution of the predicted probabilities calculated by this model (after shrinkage) for the 240 patients in this study. The individual predicted probabilities ranged from 15% to 82%. When we would consider a predicted probability at 12 months below 30% as poor prognosis, the model was able to identify 25 out of 240 (10%) as poor prognosis patients.

In an additional analysis, we defined WHO 2 patients as definitive PCOS in case they presented with elevated FAI and polycystic ovaries (Laven *et al.*, 2002) (49% of the 240 patients) - in accordance with a recent PCOS consensus meeting - and clearly non-PCOS when FAI was normal and no signs of polycystic ovaries were present (13% of patients). The remaining women formed the in-between group of possible

PCOS (38% of patients). Kaplan-Meier analysis for pregnancy leading to singleton live birth showed no difference between these groups: P = 0.9.

DISCUSSION

The current prospective study demonstrates a cumulative chance of pregnancy resulting in singleton live birth of 71% within 2 years of classical ovulation induction (CC medication followed by exogenous FSH as second line therapy) in a well-defined group of normogonadotrophic anovulatory infertile women (WHO group 2). The chance increased to 74% in case multiple live births were included, and to 76% for all ongoing pregnancies. Of all live births, only 7% were multiple births. These figures compare favorably with the outcome of alternative, more complex treatment algorithms (Heard *et al.*, 2002)-(Child *et al.*, 2001). Again, many studies involving alternative therapies are retrospective, and involve a small and selected group of patients. Moreover, singleton live birth is rarely used as an endpoint. The above mentioned observations are in favor of the contention that classical ovulation induction should remain the first line treatment of choice in these women.

A prognostic model based on age of the woman, the insulin/glucose ratio and the duration of infertility predicts probabilities of pregnancy resulting in singleton live birth at 12 months ranging from 15% to 82%. The apparent c-statistic of the model was 0.64, and the optimism-corrected value was 0.61, indicating only a moderate discriminative ability. However, the model is able to identify patients with low chances for success applying classical ovulation induction who may therefore qualify for alternative first line treatment options such as insulin sensitizing agents, assisted reproductive technologies or laparoscopic ovarian surgery. For example, a probability lower than 30% was predicted in 25 (10%) out of the 240 patients. However, this cut-off value for low chance of 30% is arbitrarily chosen and other values may be just as valid. In a general infertility setting, the individual choice between several treatment options should depend on the chances of success, the complication rates, the burden to the patient and costs. Taking these considerations into account, clinicians can choose any desired cut-off value for the chance of success of ovulation induction, and derive from Figure 3 how many patients will be identified as having a low chance of success. In previous studies by our group with regard to outcomes following CC alone, a history of amenorrhea (as opposed to oligomenorrhea) was found to be an adverse factor for ovulation (Imani *et al.*, 1998), but a positive factor for pregnancy once ovulation was achieved (Imani *et al.*, 1999). In the current study, amenorrhea has no effect on outcome, indicating that probably both effects level out, as has been noticed previously (Imani *et al.*, 2002b). Additional factors involved in the prediction of life birth following CC included FAI, BMI and female age. Previously identified screening characteristics involved in the prediction of ongoing pregnancy following FSH ovulation induction include IGF-I and T concentrations and age (Mulders *et al.*, 2003).

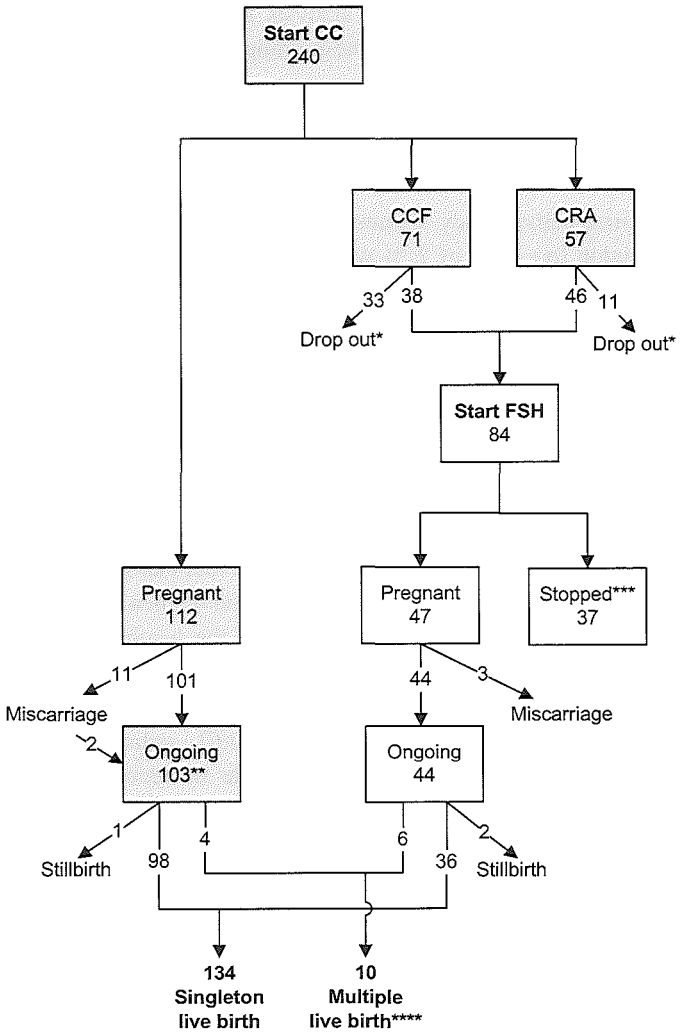


Figure 2 Outcomes during classical induction of ovulation in 240 normogonadotrophic anovulatory infertile women.

* Drop out after CC occurred in 44 patients. In 13 cases because the couple no longer had an active child wish, in 7 cases because further treatment was found too burdensome, in 9 cases because of medical problems and in 15 cases because an other therapy was started (IVF, or IUI).

** two women with miscarriages following CC presented with an ongoing pregnancy after a subsequent second (1 patient) or third (1 patient) pregnancy with CC.

*** FSH treatment was discontinued by 37 patients without getting pregnant. In 3 cases because the couple no longer had an active child wish, in 1 case because treatment was found too burdensome, in 6 cases because of medical problems and in 22 cases because the couples started another therapy (IVF). Five couples had not yet planned another treatment at the end of follow-up.

**** Multiple ongoing pregnancies with CC occurred four times, all of which were twins. With FSH, 6 multiple live birth occurred (4 twins, 1 triplet and 1 quadruplet). CCF = CC failure (absence of pregnancy despite ovulatory CC cycles. CRA = clomiphene resistant anovulation.

Table 1 Clinical, endocrine and ultrasound screening characteristics (mean ± SD) of 240 normogonadotrophic anovulatory infertile women before initiation of classical ovulation induction (CC as first line, and FSH as second line therapy), and separately for women who did and those who did not reach a pregnancy resulting in singleton live birth.

	Overall group	Pregnancy resulting in singleton live birth			
Screening parameters	(N = 240)	No (N =106)	Yes (N = 134)	P value[1]	Multivariable Hazard Ratio (95% CI) [2]
Age (yr)	27.8 ± 4.3	28.7 ± 4.5	27.1 ± 4.0	< 0.001	0.91 (0.87 - 0.95)
Duration of Infertility (yr)	2.1 ± 2.1	2.5 ± 2.5	1.9 ± 1.6	0.02	0.91 (0.81 - 1.02)
% Primary infertility (patients)	75% (n = 181)	72% (n = 76)	78% (n = 105)	0.17	-
% Oligomenorrhea (patients)	78% (n = 187)	79% (n = 84)	77% (n = 103)	0.9	-
BMI (kg/m^2)	26.9 ± 6.2	27.7 ± 6.8	26.2 ± 5.6	0.02	-
WHR	0.8 ± 0.1	0.8 ± 0.1	0.8 ± 0.1	0.28	-
FAI (Tx100/SHBG)	6.2 ± 4.6	6.3 ± 4.7	6.2 ± 4.6	0.5	-
FSH (IU/l)	4.7 ± 1.5	4.4 ± 1.4	4.8 ± 1.5	0.07	-
LH (IU/l)	7.6 ± 4.3	7.2 ± 4.0	7.9 ± 4.6	0.04	-
E_2 (pmol/l)	240 ± 136	249 ± 139	233 ± 133	0.2	-
Insulin/glucose ratio	3.4 ± 2.1	3.6 ± 2.4	3.2 ± 1.8	0.04	0.87 (0.79 - 0.96)
% Polycystic ovaries (n patients) [3]	74% (n = 178)	75% (n = 79)	74% (n = 99)	0.8	-
Total motile sperm count (TMC) [4]	68 (0 - 693)	52 (1 - 662)	80 (0 - 693)	0.9	-

[1] P value of log-Rank test

[2] The Hazard Ratios for pregnancy resulting in singleton live birth with 95% Confidence Interval (CI) of the screening characteristics that were selected into the multivariable model.

The formula (after shrinkage) to calculate the 12 month predicted probability for a new patient is:

1 - EXP[-8.3 * EXP[-0.072*age (yrs) – 0.107*(insulin/glucose ratio) – 0.069*duration(yrs)]]; 'EXP' denotes the standard exponential function.

[3] Polycystic ovaries are defined as mean ovarian volume ≥ 10.8 and/or mean follicle number per ovary ≥ 10 (van Santbrink et al. Fertil Steril 1997 (van Santbrink *et al.*, 1997)).

[4] For total motile sperm count median and range is given instead of mean and standard deviation.

Almost 50% of the CC failure (CCF) patients did not proceed to FSH treatment and should be considered as dropout. In general, dropouts may cause bias in the estimates of outcome when their prognosis differs from the patients who did not drop out ('selective dropout'), as noted in a previous study (Imani et al., 2002b). However, the screening characteristics that were predictive of success were similar between these patients and the patients who continued with FSH treatment. Only endogenous FSH levels were different, but since that screening characteristic was not associated with outcome in the prognostic model, we conclude that there is no indication of selective dropout. The 37 women who did not reach the endpoint following FSH treatment had received a median number of 5 treatment cycles, which is close to the standard number of treatment cycles according to protocol. Therefore they are not to be considered as dropouts.

It is generally perceived that the majority of patients that qualify for exogenous gonadotrophins for ovulation induction suffer from PCOS (White *et al.*, 1996), (Fauser and Van Heusden, 1997). However, scientific evidence to support this believe is scant. We have recently established that 70% of women suffering from normogonadotrophic anovulation present with polycystic ovaries as diagnosed by ultrasound, and 61% with endocrine signs associated with PCOS (i.e. hyperandrogenemia) (Laven *et al.*, 2002). However, insulin resistance – as defined by an increased fasting insulin/glucose ratio - could only be observed in 21% of women, whereas 44% were obese. All these clinical, endocrine and ultrasound abnormalities are involved in ovarian dysfunction in PCOS, and may therefore potentially predict response and outcome of ovarian stimulation by exogenous FSH (Balen *et al.*, 1993), (Fulghesu *et al.*, 1997). We were therefore surprised to find that none of these features – except for insulin resistance – could be identified in the current study as predictor of pregnancy in multivariable analysis. Moreover, in case the overall heterogeneous WHO 2 group was classified according to definitive PCOS or definitive non-PCOS (see results section for details) again no differences in pregnancy chances could be observed. Not even obesity (defined as elevated BMI) or abnormal fat distribution (defined as WHR) was capable of predicting outcome, despite much emphasis in recent literature on the impact of overweight (Hamilton-Fairley *et al.*, 1992) and weight reduction (Norman *et al.*, 2002) on success of fertility therapies in anovulatory women. These discrepancies clearly underline the need for properly designed, prospective studies in a well defined patient population.

In conclusion, the current study shows that overall normogonadotrophic anovulatory women have good prospects of singleton live birth when treated according to the classical induction of ovulation algorithm, including CC as first line treatment followed by exogenous FSH as second line. Using a multivariate prediction model, individual women with chances for success (far) above or below average can be identified. For women with a predicted low chance, the use of alternative treatment strategies may be appropriate.

Acknowledgements

This study is financially supported by 'VAZ/CVZ Doelmatigheidsonderzoek' project no 99201, and 'Stichting Voortplantingsgeneeskunde Rotterdam' (SVG).

References:

Akaike H. Information theory as an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. *International symposium on information theory*. Budapest: Akademia Kiado, 1973:267-81.

Balen AH, Tan SL, Jacobs HS. (1993) Hypersecretion of luteinising hormone: a significant cause of infertility and miscarriage. *Br J Obstet Gynaecol.* 100, 1082-9.

Buyalos RP, Lee CT. (1996) Polycystic ovary syndrome: pathophysiology and outcome with in vitro fertilization. *Fertil Steril.* 65, 1-10.

Child TJ, Abdul-Jalil AK, Gulekli B, Tan SL. (2001) In vitro maturation and fertilization of oocytes from unstimulated normal ovaries, polycystic ovaries, and women with polycystic ovary syndrome. *Fertil Steril.* 76, 936-42.

Cohen J. (1996) Laparoscopic procedures for treatment of infertility related to polycystic ovarian syndrome. *Hum Reprod Update.* 2, 337-44.

Farquhar C, Vandekerckhove P, Lilford R. (2001) Laparoscopic "drilling" by diathermy or laser for ovulation induction in anovulatory polycystic ovary syndrome. *Cochrane Database Syst Rev.* 4.

Fauser BC, Van Heusden AM. (1997) Manipulation of human ovarian function: physiological concepts and clinical consequences. *Endocr Rev.* 18, 71-106.

Fulghesu AM, Villa P, Pavone V, Guido M, Apa R, Caruso A, et al. (1997) The impact of insulin secretion on the ovarian response to exogenous gonadotropins in polycystic ovary syndrome. *J Clin Endocrinol Metab.* 82, 644-8.

Glueck CJ, Wang P, Goldenberg N, Sieve-Smith L. (2002) Pregnancy outcomes among women with polycystic ovary syndrome treated with metformin. *Hum Reprod.* 17, 2858-64.

Hamilton-Fairley D, Kiddy D, Watson H, Paterson C, Franks S. (1992) Association of moderate obesity with a poor pregnancy outcome in women with polycystic ovary syndrome treated with low dose gonadotrophin. *Br J Obstet Gynaecol.* 99, 128-31.

Hammond MG, Halme JK, Talbert LM. (1983) Factors affecting the pregnancy rate in clomiphene citrate induction of ovulation. *Obstet Gynecol.* 62, 196-202.

Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA. (1985) Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep.* 69, 1071-77.

Harrell FE, Jr., Lee KL, Mark DB. (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 15, 361-87.

Heard MJ, Pierce A, Carson SA, Buster JE. (2002) Pregnancies following use of metformin for ovulation induction in patients with polycystic ovary syndrome. *Fertil Steril.* 77, 669-73.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (1998) Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 83, 2361-5.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (1999) Predictors of chances to conceive in ovulatory patients during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 84, 1617-22.

Imani B, Eijkemans MJ, de Jong FH, Payne NN, Bouchard P, Giudice LC, et al. (2000) Free androgen index and leptin are the most prominent endocrine predictors of ovarian response during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 85, 676-82.

Imani B, Eijkemans MJ, Faessen GH, Bouchard P, Giudice LC, Fauser BC. (2002a) Prediction of the individual follicle-stimulating hormone threshold for gonadotropin induction of ovulation in

normogonadotropic anovulatory infertility: an approach to increase safety and efficiency. *Fertil Steril.* 77, 83-90.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (2002b) A nomogram to predict the probability of live birth after clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *Fertil Steril.* 77, 91-7.

Kousta E, White DM, Franks S. (1997) Modern use of clomiphene citrate in induction of ovulation. *Hum Reprod Update.* 3, 359-65.

Laven JS, Imani B, Eijkemans MJ, Fauser BC. (2002) New Approach to Polycystic Ovary Syndrome and Other Forms of Anovulatory Infertility. *Obstet Gynecol Surv.* 57, 755-767.

Legro RS. (1998) Polycystic ovary syndrome: current and future treatment paradigms. *Am J Obstet Gynecol.* 179, S101-S108.

Little RJA. (1992) Regression with missing X's: a review. *J Am Stat Assoc.* 87, 1227-1237.

Messinis IE, Milingos SD. (1997) Current and future status of ovulation induction in polycystic ovary syndrome. *Hum Reprod Update.* 3, 235-53.

Mitwally MF, Casper RF. (2001) Use of an aromatase inhibitor for induction of ovulation in patients with an inadequate response to clomiphene citrate. *Fertil Steril.* 75, 305-9.

Mulders AG, Eijkemans MJ, Imani B, Fauser BC. (2003) Prediction of Chances for Success or Complications in Gonadotropin Ovulation Induction in Normogonadotropic Anovulatory Infertility. *Reprod Biomed Online.* 7, In Press.

Nestler JE, Stovall D, Akhter N, Iuorno MJ, Jakubowicz DJ. (2002) Strategies for the use of insulin-sensitizing drugs to treat infertility in women with polycystic ovary syndrome. *Fertil Steril.* 77, 209-15.

Norman RJ, Davies MJ, Lord J, Moran LJ. (2002) The role of lifestyle modification in polycystic ovary syndrome. *Trends Endocrinol Metab.* 13, 251-7.

Pache TD, Wladimiroff JW, Hop WC, Fauser BC. (1992) How to discriminate between normal and polycystic ovaries: transvaginal US study. *Radiology.* 183, 421-3.

The ESHRE Capri Workshop Group. (1995) Anovulatory infertility. *Hum Reprod.* 10, 1549-53.

van Houwelingen JC, Le Cessie S. (1990) Predictive value of statistical models. *Stat Med.* 9, 1303-25.

van Santbrink EJ, Fauser BC. (1997) Urinary follicle-stimulating hormone for normogonadotropic clomiphene- resistant anovulatory infertility: prospective, randomized comparison between low dose step-up and step-down dose regimens. *J Clin Endocrinol Metab.* 82, 3597-602.

van Santbrink EJ, Hop WC, Fauser BC. (1997) Classification of normogonadotropic infertility: polycystic ovaries diagnosed by ultrasound versus endocrine characteristics of polycystic ovary syndrome. *Fertil Steril.* 67, 452-8.

White DM, Polson DW, Kiddy D, Sagle P, Watson H, Gilling-Smith C, et al. (1996) Induction of ovulation with low-dose gonadotropins in polycystic ovary syndrome: an analysis of 109 pregnancies in 225 women. *J Clin Endocrinol Metab.* 81, 3821-4.

11

Individualizing the treatment protocol

The prediction models discussed so far in part II of this thesis may serve several purposes: they a) may be used to inform patients about the prospects of the outcome of treatment, b) serve as a basis of decision making about which course of treatment is to be preferred for an individual patient and finally, c) be used to individualize and optimize the treatment protocol. For the first purpose (a), it suffices the have an easy to use tool that will allow the clinician to calculate the prognosis of the patients, given their characteristics. Examples of such tools are a score chart, such as the one used in *Chapter 7* for the prediction of the chance to be CRA, or a nomogram, such as the one presented in *Chapter 9* to calculate the chance of pregnancy leading to live birth with CC treatment. The second purpose (b) involves knowledge on the individual preferences of patients for the possible outcomes that may result from treatment. These preferences may concern the time required to achieve a pregnancy, the burden of treatment or whether there is a chance of a twin or higher order multiple pregnancy. From literature we know that preferences are difficult to assess and the values obtained from individual patients are very unreliable (Roest *et al.*, 1997). The third purpose (c) requires the calculation of the cumulative pregnancy chances for various alternative treatment protocols and relating these to the costs, for patient groups defined by the predictive variables. The result will be a new treatment protocol that is individualized and optimized with respect to effects and costs. Of the three purposes mentioned this is the one most likely to be of practical value, and it will be pursued here.

EFFICIENCY PROBLEM

The current sequence of treatment is from the least burdensome and cheapest to the most burdensome and most costly treatment. The question is whether the sequence of treatments can be made more cost-effective by using information on individual patient characteristics and the response of the patient in previous treatment steps.
Our aim is to investigate for which patient groups the current linear protocol may be replaced with a protocol in which:

1. The CC step, the GOI-step or both may be skipped, based on the patient's characteristics and (for the second option) on findings during the CC treatment (was the patient CRA or a Clomiphene Failure; what was the CC-dose at first ovulation?)
2. Further treatment is stopped, based on the patient's characteristics and findings during treatments so far, or no treatment is started because the spontaneous pregnancy chances of the patient are still good enough.

The identification of patient groups for whom a change of treatment protocol would be cost-effective is based on the prediction models for the chance of response (ongoing pregnancy) and an estimate of the costs of treatment.

PREDICTION OF PREGNANCY CHANCES

So far in this thesis, prediction models have been presented for the chance of pregnancy with CC (*Chapters 7, 8* and *9*) and the overall prognosis of ovulation induction (CC + GOI) in *Chapter 10*. For the cost-effectiveness analysis, additional prediction models are needed for the GOI step separately, for the outcome of IVF treatment as well as for the chances to become pregnant spontaneously.

Prediction of outcome of GOI

A separate study has been performed to obtain a prediction model for the outcome after GOI (Mulders *et al.*, 2003a). For this purpose 154 patients were included that all had a history of CC treatment either ending with CRA of Clomiphene Failure. There were 67 (= 44% cumulative) ongoing pregnancies. A multivariable Cox regression model was constructed based on the initial serum insulin-like growth factor-I (IGF-I), testosterone and the woman's age. The response during the CC treatment (CRA versus Clomiphene Failure) was not associated with the cumulative chance of ongoing pregnancy (P = 0.6).

Prediction of outcome of IVF

From the Rotterdam cohort of 240 patients that was analyzed in *Chapter 10*, only 26 patients started IVF after having received foregoing CC and GOI treatments without success (Mulders *et al.*, 2003c). These patients were compared with a control group of IVF patients with tubal pathology. Compared to the controls, women with WHO II anovulatory infertility had an elevated risk of cancellation of an IVF treatment cycle because of inadequate ovarian response to stimulation, but the pregnancy rates were similar. No predictors for pregnancy after IVF could be identified within this small group of patients. We therefore based our calculations of the prognosis with IVF on the Templeton model (Templeton *et al.*, 1996). This model was built on the data from the national IVF registry in the United Kingdom, and is based on a very large number of patients from the period 1991-'94. The overall livebirth rate in the first cycle was 14%. Since the results of IVF may have improved over time due to technological developments, we corrected the baseline pregnancy rate of this model with recent published data from the national Dutch IVF registration (Kremer *et al.*, 2002), which had an ongoing pregnancy rate per cycle of 19% during the period 1998-2000. Pregnancy chances with IVF were subsequently calculated for ages under and above 30 years as 24% and 19% respectively per treatment cycle. Assuming that on average 2.1 cycles will be performed (Kremer *et al.*, 2002), the overall pregnancy chances with IVF are 50% and 40% respectively.

Prediction of spontaneous pregnancy chances

Although some patients became pregnant spontaneously before as well as between treatments, we do not have reliable estimates of the chances on spontaneous pregnancy for our own patient group. Fortunately, predictive models for spontaneous pregnancy chances have been published, including prognosis for patients with ovulation disorders (Collins *et al.*, 1995), (Snick *et al.*, 1997). For our analysis we will use the Snick model (Snick *et al.*, 1997), since this model was developed on patients who were directly referred by general practitioners, which is also the case for our patient group. The Snick model provides estimates dependent on *age, duration of infertility, type of infertility (primary or secondary)* and *diagnosis*. We will assume for the WHO II patients that the duration of infertility is shorter than 2 years and that infertility is primary. For women with an ovulation disorder and age under 30, the model predicts a spontaneous pregnancy chance within one year of 20%. However, we may assume that amenorrheic women do not ovulate, and therefore have a zero chance of spontaneous pregnancy. This implies that oligomenorrheic women must have a pregnancy chance that is above the average for women with an ovulation disorder. In our group of 240 patients, 22% had amenorrhea. Therefore, the following equation should hold: 20% = 22% x 0 + 78% x P(pregnant | oligomenorrhea). From this equation follows that the spontaneous pregnancy chance for oligomenorrheic women, younger than 30 may be estimated as being 0,20/0,78 = 26%. For women of 30 years or older the calculation is

similar: the Snick model predicts 15% for the whole group of women with an ovulation disorder. In case of oligomenorrhea, the prognosis is then equal to 0,15/0,78 = 19%.

Summary of the predictors

The prediction models for outcome of CC presented in *Chapters 7, 8* and *9* (Imani *et al.*, 1998), (Imani *et al.*, 1999), (Imani *et al.*, 2002), models for GOI and IVF (Mulders *et al.*, 2003a), (Mulders *et al.*, 2003c), the findings from literature concerning the outcome of GOI (Mulders *et al.*, 2003b) and the results from the 'overall' analysis for ovulation induction (CC + GOI) in *Chapter 10* (Eijkemans *et al.*, 2003) show that body mass index (BMI), female age, the severity of cycle disturbance (oligo- or amenorrhea), serum androgen levels and ultrasound characteristics of the ovaries are related to the outcomes of treatment. **Table 1** gives a resume of the predictors that play a role in the treatment of these patients.

Table 1: Predictors for chances on outcome of the three treatment modalities

Predictor	CRA[a]	Pregnancy with CC in non- CRA patients[b]	Pregnancy with GOI[c]	Pregnancy with IVF[d]	Spontaneous pregnancy[e]
Age		x	x	x	x
BMI	x				
Oligo/amenorrhea	x	x			x
Androgens	x		x		
Ovaries	x				

[a] (Imani *et al.*, 1998)
[b] (Imani *et al.*, 1999)
[c] (Mulders *et al.*, 2003a): this study also found that IGF-I was predictive of pregnancy chances with GOI. Since IGF-I is not part of the standard initial hormonal assessment, we decided to leave it out of our analysis.
[d] (Templeton *et al.*, 1996)
[e] (Snick *et al.*, 1997)

COSTS OF TREATMENT

The costs of the three treatment options were calculated from a societal perspective, as the product of volumes and prices. On the data of the 240 patients that were used in the overall analysis in *Chapter 10* (Eijkemans *et al.*, 2003), the average costs per treatment cycle were calculated. The total costs of treatment depend on the number of treatment cycles, which is related to the prognosis: the higher the pregnancy chances, the sooner the patient will become pregnant, needing less cycles. Therefore, we will

present the costs per cycle separately and postpone the calculation of the total costs to the final cost-effectiveness analysis.

Costs per cycle of CC treatment

In calculating the costs of a CC treatment cycle, we had to make a number of simplifying assumptions on the amount of medication and tests used during one cycle of CC treatment. These assumptions are shown in **Table 2**, together with the prices per unit.

The dose escalation only applies to patients who do not ovulate with the previous dose, and will occur two times in patients who turn out to be CRA. Therefore, the costs per cycle are higher for CRA patients than for non-CRA patients: they amount to € 30,76 and € 16,10 respectively. The average costs per cycle for all patients were € 18,96.

Table 2: Prices and assumptions on the volume per cycle of CC treatment

	Price	Volume per cycle
Medication Clomid	10 tablets of 50 mg: € 4,62	Each cycle 5 days, dose escalating from 50mg/day in the first cycle, to max 150 mg/day in subsequent cycles
Progesterone assessment	€ 25	1 x first cycle, in subsequent cycle only when the daily dose was increased
Pregnancy test	€ 10	Average 1x per 2 cycles

Costs per cycle of GOI treatment

In calculating the costs of a GOI treatment cycle we made the assumptions mentioned in **Table 3**.

In total, the costs of ovulation induction with gonadotrophins were € 785 per cycle, which is more than 40 times as high as the costs per CC-cycle.

Table 3: Prices and assumptions on the volume per cycle of GOI treatment

	Price	Volume per cycle
Medication Recombinant FSH, per ampoule 75 IE	€ 33,54	Mean number of ampoules: 17,7 per cycle
Echo	€ 40	Mean: 4,4 per cycle
Injection hCG, 5000 IE	€ 4,63	Only in non-cancelled cycles.
Pregnancy test	€ 10	Every cycle

Costs per cycle of IVF treatment

For the costs of IVF we used data from the literature, because our own data were too limited to give reliable estimates. The costs of IVF in the Netherlands have been determined in the study by Goverde et al. (Goverde *et al.*, 2000). That study gave an estimate of the costs per IVF cycle of DFL 3.350 (price level of 1995), i.e. € 1.529, about twice as high as the price per cycle of GOI.

OPTIMIZATION OF THE TREATMENT PROTOCOL BASED ON THE COSTS PER ONGOING PREGNANCY

Here we present the results of the cost-effectiveness calculations. First we will define patient groups according to combinations of the most important predictor variables. Next we determine –for every patient group- the total costs and ongoing pregnancy rate of each treatment. Then the treatment strategies will be compared in cost-effectiveness, and finally alternatives will be discussed.

Predictors and patient groups

We have seen that body mass index (BMI), female age, the severity of cycle disturbance (oligo- or amenorrhea), serum androgen levels and ultrasound characteristics of the ovaries were predictive of outcome of treatment. An individualized treatment protocol based on all these characteristics would be too complicated. Therefore, we restricted the number of predictors to be used in the cost-effectiveness analysis to the 4 most important ones: age, the severity of cycle disturbances (oligo- or amenorrhea), androgen levels expressed as normal or elevated (Androstenedione >=16.3 nmol/L ór testosterone >= 3.2 nmol/L), and Body Mass Index (BMI). Age and BMI were dichotomized, so that 16 (=2x2x2x2) different patient groups resulted. The patient groups are presented in **Table 4** together with their frequency of occurrence in the Rotterdam cohort of 240 patients.

The prediction models for the chance to be CRA, the pregnancy chance with CC in non-CRA patients and the pregnancy chance with GOI were re-estimated with the dichotomized predictors as defined here. Shrinkage of regression coefficients was applied, just as it had been done in the original models.

Comparison of treatment strategies

The costs and pregnancy chances of the three treatments were calculated for the 16 patient groups. Next, these results were incorporated in the treatment strategies that we want to compare. We performed a cost effectiveness analysis of a strategy in which the CC step is skipped and treatment starts with GOI followed by IVF (GOI + IVF), and a strategy in which the GOI step is skipped and treatment after unsuccessful CC is IVF (CC + IVF). These two alternative treatment strategies will be compared to the

Table 4: Definition of the 16 patient groups

Age	Cycle duration	Androgens	BMI	% of patients
Age < 30	Oligomenorrhea	Normal	BMI < 27	18.8%
			BMI ≥ 27	13.3%
		Elevated	BMI < 27	10.0%
			BMI ≥ 27	10.0%
	Amenorrhea	Normal	BMI < 27	5.4%
			BMI ≥ 27	2.9%
		Elevated	BMI < 27	5.0%
			BMI ≥ 27	3.3%
Age ≥ 30	Oligomenorrhea	Normal	BMI < 27	8.8%
			BMI ≥ 27	9.2%
		Elevated	BMI < 27	5.0%
			BMI ≥ 27	2.9%
	Amenorrhea	Normal	BMI < 27	2.5%
			BMI ≥ 27	0.8%
		Elevated	BMI < 27	1.3%
			BMI ≥ 27	0.8%

reference strategy CC + GOI + IVF. Next to this, the decision to stop further treatment, or even not commence with treatment, will be analyzed by comparing the prognosis after treatment with the spontaneous pregnancy prognosis.

The costs and the ongoing pregnancy chances of the various treatment strategies are compared to those of the reference strategy, and the incremental cost-effectiveness ratio (extra costs per extra ongoing pregnancy) will be calculated. In case the costs are higher and at the same time the pregnancy chances are lower, the alternative strategy is unacceptable (it is dominated by the reference strategy). In the reverse case, with lower costs and higher ongoing pregnancy chances we have a 'win-win' situation (the alternative dominates the reference strategy). In the two other possibilities (higher costs together with higher ongoing pregnancy chances, or when both are lower) the height of the cost-effectiveness ratio determines whether the alternative is acceptable. If the cost-effectiveness ratio is below the threshold-value of € 10.000 per extra ongoing pregnancy, we consider the alternative acceptable. We also explored a lower and a higher value for the threshold cost-effectiveness ratio, of € 5.000 and € 20.000 respectively.

For the comparison of the alternative treatment strategies with the strategy CC + GOI + IVF we have to make a few additional assumptions. The flowchart in *Chapter 10* showed that not all patients continue with GOI after unsuccessful CC treatment. We have seen that 81% of the CRA patients, but only 54% of the Clomiphene Failures, continued treatment with GOI. Further, only 59% of the patients who were

not pregnant after GOI continued with IVF. We will assume that in the alternative treatment strategies the same percentage of non-pregnant patients move to the next treatment step after CC resp. GOI.

The results of the cost-effectiveness analysis are presented in Table 5. The treatment strategy GOI + IVF (i.e. skip CC) is either dominated by the reference strategy or produces more ongoing pregnancies against costs per pregnancy that vary between € 30.000 and € 70.000, which is above threshold value. Therefore this strategy would not be acceptable for any patient group. Under the lower as wel as under the higher threshold value, the conclusion remains the same.

The other alternative treatment strategy, CC + IVF (i.e. skip GOI), produces less pregnancies and lower costs than the reference strategy. Therefore, the cost-effectiveness ratio represents the extra costs per extra ongoing pregnancy of the reference strategy compared to the strategy CC+IVF. The ratio is within the threshold for acceptability of € 10.000, except for women over 30 years of age who have elevated androgen levels. For them, GOI after CC hardly gives extra ongoing pregnancies but does increase the costs substantially, with a cost-effectiveness ratio of € 96.000. This concerns about 14% of the group of patients that is not pregnant after CC and who would start GOI treatment.

Other alternatives

Apart from skipping treatment steps, one could also consider stopping treatment in case of very poor prognosis, or not start at all in case the spontaneous pregnancy chances are still high enough. These options may be evaluated by calculating the extra costs per extra ongoing pregnancy of continuing treatment compared to trying to become pregnant spontaneously. We make the simplifying assumption that trying to become pregnant spontaneously doesn't require extra costs.

The extra costs per extra ongoing pregnancy of the reference strategy (CC+GOI+IVF) compared to the spontaneous pregnancy chances vary between € 1,260 and € 8,500. Assuming that spontaneous pregnancy chances are unaffected after unsuccessful CC treatment, the cost-effectiveness ratio of continuing with GOI + IVF after CC compared to trying to become pregnant spontaneously varies between € 4,840 and € 20,260 per pregnancy. Finally, the cost-effectiveness ratio of continuing with IVF after unsuccessful GOI compared to trying to become pregnant spontaneously varies between € 6,420 and € 15,290. The highest values suggest that indeed there are patient groups for whom continuing treatment is not efficient under the threshold value of € 10.000 per extra ongoing pregnancy. However, we note that it is likely that the spontaneous pregnancy chances of patients who didn't become pregnant with CC or GOI are lower than assumed here (a.o. because the couples with better chances already will have become pregnant). Therefore, the cost-effectiveness ratios are probably lower than we have calculated and we cannot conclude for any patient group that stopping treatment is efficient.

Table 5 Costs and effects of two alternative treatment strategies (GOI + IVF and CC + IVF) compared to the reference strategy CC + GOI + IVF

				Reference strategy: CC+GOI+IVF			Alternative 1: GOI+IVF			Alternative 2: CC+IVF		
Age	Cycle duration	Androgens	BMI	Pregnancy chance	Costs	Costs/ preg	Pregnancy chance	Costs	C/E ratio*	Pregnancy chance	Costs	C/E ratio*
Age < 30	Oligomenorrhea	Normal	BMI < 27	0.76	808	1,064	0.71	3,422	-**	0.72	763	1,023
			BMI ≥ 27	0.73	947	1,291	0.71	3,422	-	0.68	893	1,023
		Elevated	BMI < 27	0.71	1,074	1,521	0.61	3,877	-	0.68	903	6,315
			BMI ≥ 27	0.64	1,473	2,289	0.61	3,877	-	0.61	1,235	6,315
	Amenorrhea	Normal	BMI < 27	0.75	944	1,261	0.71	3,422	-	0.7	891	1,023
			BMI ≥ 27	0.68	1,419	2,092	0.71	3,422	70,284	0.6	1,337	1,023
		Elevated	BMI < 27	0.63	1,626	2,562	0.61	3,877	-	0.59	1,362	6,315
			BMI ≥ 27	0.55	2,351	4,237	0.61	3,877	30,126	0.49	1,964	6,315
Age ≥ 30	Oligomenorrhea	Normal	BMI < 27	0.54	1,533	2,844	0.5	4,160	-	0.5	1,203	9,132
			BMI ≥ 27	0.52	1,673	3,195	0.5	4,160	-	0.48	1,311	9,132
		Elevated	BMI < 27	0.49	1,826	3,742	0.41	4,533	-	0.48	1,318	96,057
			BMI ≥ 27	0.45	2,201	4,916	0.41	4,533	-	0.44	1,586	96,057
	Amenorrhea	Normal	BMI < 27	0.56	1,630	2,931	0.5	4,160	-	0.52	1,278	9,132
			BMI ≥ 27	0.5	2,097	4,160	0.5	4,160	-	0.45	1,640	9,132
		Elevated	BMI < 27	0.46	2,305	5,038	0.41	4,533	-	0.45	1,661	96,057
			BMI ≥ 27	0.39	2,971	7,550	0.41	4,533	77,269	0.38	2,135	96,057
			Average	0.66	1,438	2,165	0.67	3,915	287,937	0.59	1,194	3,454

* C/E ratio: the extra costs per extra ongoing pregnancy of the treatment strategy compared with the reference strategy CC + GOI + IVF.

** With "-" the cases are indicated in which the strategy is dominated by the reference strategy (i.e. the strategy has higher costs and lower pregnancy chances than the reference strategy).

CONCLUSIONS

For most patients, the current standard treatment protocol is cost-effective. An exception may be made for women above 30 years of age with elevated androgen levels, for whom the treatment with gonadotrophins could be skipped.
The strong association between outcome of treatment and obesity (for CC from our own analysis, for gonadotrophins mainly from literature) implies that chances could be improved by reducing overweight. This concerns 30-40% of the WHO II group.

References

Collins JA, Burrows EA, Wilan AR. (1995) The prognosis for live birth among untreated infertile couples. *Fertil Steril.* 64, 22-8.

Eijkemans MJ, Imani B, Mulders AG, Habbema JD, Fauser BC. (2003) High singleton live birth rate following classical ovulation induction in normogonadotrophic anovulatory infertility (WHO 2). *Hum Reprod.* 18, 2357-62.

Goverde AJ, McDonnell J, Vermeiden JP, Schats R, Rutten FF, Schoemaker J. (2000) Intrauterine insemination or in-vitro fertilisation in idiopathic subfertility and male subfertility: a randomised trial and cost-effectiveness analysis. *Lancet.* 355, 13-8.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (1998) Predictors of patients remaining anovulatory during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 83, 2361-5.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (1999) Predictors of chances to conceive in ovulatory patients during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *J Clin Endocrinol Metab.* 84, 1617-22.

Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. (2002) A nomogram to predict the probability of live birth after clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrheic infertility. *Fertil Steril.* 77, 91-7.

Kremer JA, Beekhuizen W, Bots RS, Braat DD, van Dop PA, Jansen CA, et al. (2002) Resultaten van in-vitro fertilisatie in Nederland, 1996-2000. *Ned Tijdschr Geneeskd.*, 146, 2358-63.

Mulders AG, Eijkemans MJ, Imani B, Fauser BC. (2003a) Prediction of Chances for Success or Complications in Gonadotropin Ovulation Induction in Normogonadotropic Anovulatory Infertility. *Reprod Biomed Online.* 7, In Press.

Mulders AG, Laven JS, Eijkemans MJ, Hughes EG, Fauser BC. (2003b) Patient predictors for outcome of gonadotrophin ovulation induction in women with normogonadotrophic anovulatory infertility: a meta-analysis. *Hum Reprod Update.* 9, 429-49.

Mulders AG, Laven JS, Imani B, Eijkemans MJ, Fauser BC. (2003c) IVF outcome in anovulatory infertility (WHO group 2) - including polycystic ovary syndrome - following previous unsuccessful ovulation induction. *Reprod Biomed Online.* 7, 50-58.

Roest FH, Eijkemans MJ, van der Donk J, Levendag PC, Meeuwis CA, Schmitz PI, et al. (1997) The use of confidence intervals for individual utilities: limits to formal decision analysis for treatment choice. *Med Decis Making.* 17, 285-91.

Snick HK, Snick TS, Evers JL, Collins JA. (1997) The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. *Hum Reprod.* 12, 1582-8.

Templeton A, Morris JK, Parslow W. (1996) Factors that affect outcome of in-vitro fertilisation treatment. *Lancet.* 348, 1402-6.

12

Discussion and conclusions part I

The first part of this thesis addresses two research questions from clinical reproductive medicine in a large historical natural fertility population. Here, we will summarize answers to these questions and discuss their implications for clinical practice. Next, specific aspects of natural fertility in the study population and consequences for interpretation of some of the results will be considered. Finally, a discussion is presented on the opportunities and limitations in general of historical demographic data to study research questions from clinical reproductive medicine, and topics for further research are proposed.

ANSWERS TO THE RESEARCH QUESTIONS

1. **What is the variation in female reproductive life span? Is the reproductive life span correlated to fertility at young age and if so, how strong is this correlation?**

The median age at last childbirth was 41 years, which is compatible with the values found in many other natural populations. The variation is large: 80% of the women had their last child between ages 33 and 45. The estimates of the end of effective fertility -derived from the distribution of age at last childbirth- at ages 20, 25, 30, 35 and 40 were 5%, 8%, 11%, 19% and 47% respectively. A part of this variation is explained by the fertility at young age. Women with higher fertility at young age showed a later end of reproductive capacity. This finding is in line with the theory of ovarian aging.

2. **Are fertility chances after 12 months unsuccessful trying to conceive small, as implied by the current standard of calling a couple infertile after 12 months?**

After 12 months of primary infertility in patients under 35 years of age, and 12-24 months of secondary infertility irrespective of age, prospects are on average still good, since 30% or more of patients will become pregnant within the next year. For young

women (age < 30) with secondary infertility prospects are still good even after 3 years of infertility.

IMPLICATIONS FOR CLINICAL REPRODUCTIVE MEDICINE

The studies described in this part of the thesis give an indication of how many women are still able to have a live birth at the relatively late ages at which women nowadays start their reproductive career. Clinicians and policy makers are concerned about the impact on health care resources of further postponement of first childbirths by women in the Netherlands and elsewhere. Individual women are concerned about how long they can wait before their chances of becoming pregnant and have a live birth, deteriorate.

The results in *chapter 2* tell us something about the probability that a woman who starts trying to get children at a certain age will not succeed to do so by natural means. These results are mainly interesting for individual couples. The results in *chapter 3* show the age- and duration-dependence of the chance to conceive a pregnancy leading to a live birth. In women under 35, after one year of primary infertility or 1-2 years of secondary infertility, prospects are good. If diagnostics are started at 1-2 years duration of infertility (earlier only on indication) and nothing is found, prognosis is better than the population average at that duration of infertility, and therapy could be withheld for the next 1 or 2 years.

These results may be combined with the present pattern of the age at which women start to get children in the Netherlands to predict the demand for clinical fertility services. This may be done for the usual definition of infertility, i.e. after one year of

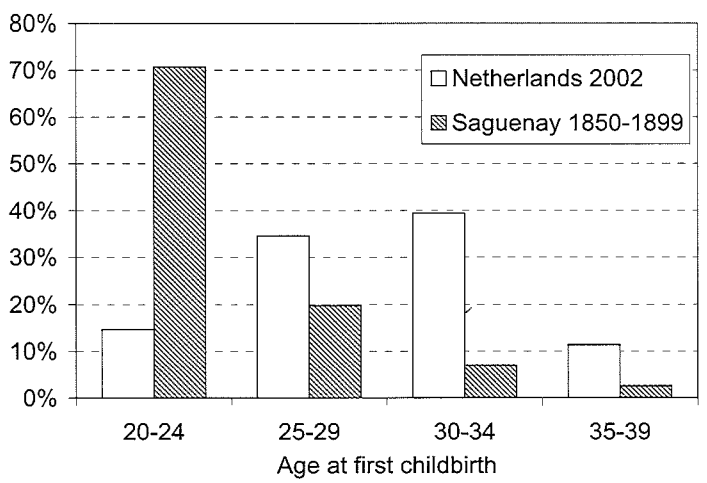


Figure 1: Distribution of the woman's age at the time of first childbirth in the Netherlands in 2002 compared to that of the women in the study population.

unsuccessful attempts to become pregnant, but also when a shorter or longer period is used to define infertility as a clinical problem. In **Figure 1**, the distribution of the woman's age at the time of first childbirth in the Netherlands in 2002 (Netherlands Central Bureau of Statistics, *www.cbs.nl*) is compared to that of the women in the study population. It is evident that women in the Netherlands nowadays have their first child at a much higher age than the women born between 1844 and 1899 in the SLSJ region.

In **Table 1,** the estimates from the analysis in *Chapter 3* are used to predict the percentages of women who are not pregnant of their first child after three different durations that could be used to define infertility: the usual duration of 12 months, a shorter duration of 9 months and a longer duration of 18 months.

Table 1: Percentages of women who are not pregnant of their first child at 9, 12 or 18 months after marriage. Women born between 1844 and 1899 in the SLSJ region

	% of women not pregnant at		
Age at marriage	9 months	12 months	18 months
20-24	18%	14%	10%
25-29	26%	21%	16%
30-34	31%	28%	20%
35-39	48%	46%	32%

These results are combined with the distribution of age at first childbirth in **Figure 1** to predict the number of women who would not yet be pregnant of their first child in the Netherlands in 2002.

Table 2: Number of women who would not yet be pregnant in the Netherlands 2002, according to their distribution of age at first childbirth (Figure 1) and the predicted percentages per age category from Table 1.

	Number of women (x 1,000) not pregnant at		
Age at marriage	9 months	12 months	18 months
20-24	2.4	1.8	1.3
25-29	8.0	6.3	5.0
30-34	10.8	9.7	6.9
35-39	4.8	4.6	3.2
Total	26.0	22.4	16.4

The total number of women not pregnant of their first child at 9 or at 18 months represent an increase of 16% or a decrease with 27% compared to the numbers not

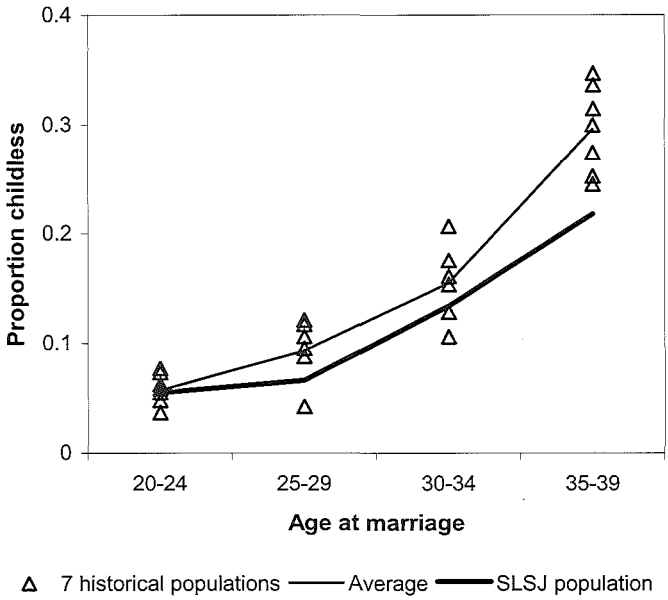


Figure 2: The proportion childlessness according to age at marriage. Seven historical populations (Menken *et al.*, 1986), compared to the women from the SLSJ, born between 1844 and 1899. Data on complete marriages.

pregnant at 12 months. This would also be the change in demand for fertility services when the standard period of 12 months to define infertility is changed to 9 or 18 months respectively. Note that we used only the data on first childbirths, therefore this calculation only applies to primary infertility. The same exercise could be performed for secondary infertility.

ASPECTS OF NATURAL FERTILITY IN THE STUDY POPULATION

Comparison with other natural fertility populations

The number of children per women in the SLSJ population was high in comparison with other historical populations. There are several explanations for this: The age at marriage was young compared to historical European populations, and there was a tradition of breastfeeding children for a relatively short period. Therefore, married women had both a long time available to produce children and did so at a relatively high rate.

The proportion of women that remained childless in the SLSJ population was 7.4% in the whole population and 6.2% in the group of women who married between ages

20 and 24 (see *chapter 3*). When comparing these figures with other demographic data, care must be taken that this comparison is not disturbed by differences between the populations in mortality: Women or men who died early, without having children at that moment, would not necessarily have remained childless, when they had lived longer. In *chapter 3*, we dealt with this by considering these couples as censored at the time of death in the Kaplan-Meier method. In demography, a different approach is taken by selecting complete families (i.e. unions still existing at age 50). With this approach, we find that the proportion childlessness in our population is 7.6% and 5.5% in the age group 20-24. Demographers have reported values as high as 8 to 10% for this age-group (Vincent, 1950) (Henry, 1965) (Trussell and Wilson, 1985) (Barrett, 1986) (Menken *et al.*, 1986). In **Figure 2**, the proportion childlessness–according to age at marriage- in our population is compared to the data on 7 populations from the 18th and 19th century, mostly from Europe (Menken *et al.*, 1986). It is seen that our population is in the lower end of the rather large variation. However, a few other populations have lower proportions of women remaining childless. Data on women from the Hutterite sect, analysed by Tietze (Tietze, 1957) showed 2.4% childlessness in women married before the age of 25. In a sub-selection of the data collected by Henry (Henry, 1965), from historical France (Kindly made available by prof. Henri Leridon), 3.7% of the women married before their 25th birthday remained childless. This subgroup consisted of women married between 1670 and 1790.

In *chapter 3*, the proportion childlessness was estimated by the Kaplan-Meier method on all marriages (censoring for early ending marriages). The estimates were 6.2% and 7.7% in the age-at-marriage categories 20-24 and 25-29 respectively. These estimates should be equal to the proportion estimated in the completed families -where we find 5.5% and 6.6%- unless there is selection in the latter on fertility (corresponding to informative censoring in the Kaplan-Meier analysis). Since the Kaplan-Meier estimates are 0.7% and 1.1% higher than the estimates on complete families, informative censoring cannot be excluded.

Comparison with modern times

We established above that the proportion childlessness in the study population falls within the range that was found in other natural populations. However, the aim of our studies was to learn lessons from the past for Reproductive Medicine, not to compare fertility in one historical population to that found in others. We therefore have to answer the question: Is infertility in a modern western society comparable to that in a 19th century rural society? In *chapter 3* we assumed that the proportion childlessness in a modern society is about half the proportion observed in this population (7.4%). This assumption is based on the literature: In 1990, Greenhall and Vessey published a review of the literature together with data from their own study on the prevalence of subfertility (Greenhall and Vessey, 1990). Studies using interviews or postal questionnaires gave rates for unresolved primary subfertility of

between 2.4% and 5.9%. Below we will discuss possible reasons for the relatively high proportion of childlessness in historical populations.

Possible reasons for the high incidence of childlessness

There is an intriguing disparity between two aspects of fertility in the study population: a relatively high incidence of childlessness compared to modern times, together with a very high number of children per woman. We discuss several possible reasons for the high incidence of childlessness, some of which are particular for our population, while others apply to all historical populations.

High incidence of genetic disorders?

This population is known to have a high incidence for several genetic disorders, probably due to a founder effect (De Braekeleer and Gauthier, 1996). None of the studies done so far revealed a difference in fertility between affected and non-affected men and women (Roy *et al.*, 1989) (Veillette *et al.*, 1989) (De Braekeleer *et al.*, 1990) (Dao *et al.*, 1992) (De Braekeleer, 1993). Further, there doesn't seem to be a biological explanation for fertility being linked to any of these diseases.

Emigration?

Because we needed information about the end of marriage, we only used the data of women who died in the region, i.e. women who did not emigrate. We stated in *chapter 1* that emigration may pose a problem: if women who had children were more likely to emigrate than women without children, the women who did not emigrate are not representative of the total group of married women: they will have a higher proportion of childlessness. In *chapter 1* we found –for women who had a child- that there was no difference in time to first pregnancy between emigrating women and women who stayed in the region. However, this does not exclude the possibility that women without children are over-represented in the group that did not emigrate.

Voluntary childlessness?

Finally, we cannot exclude that a substantial part of the high proportion childlessness is voluntary, i.e. through the use of contraceptive measures. Then it would also be likely that there were couples that used contraceptives to limit their family size. However, in order to achieve the very high average number of children that we observe, a part of the population must have had a very high fertility. It would be interesting to look at the distribution of births in completed families and compare it with other populations. It is unlikely that this factor can explain why in almost all historical populations the infertility (measured by the proportion childlessness) was higher than it is today.

Causes of infertility

Another issue is whether the causes of infertility in our study population are the same as in modern times. There are speculations that infertility is higher nowadays than it used to be, because of two trends. Tubal disease may have become more frequent during the twentieth century, as a consequence of the increased incidence of Pelvic Inflammatory Disease (PID). The primary cause of PID is sexually transmitted diseases (STD's) such as gonorrhea and chlamydia (Jones *et al.*, 1982) (Kane *et al.*, 1984). Indeed the incidence of STD's increased in the USA between 1965 and 1975, but remained constant thereafter (Curran, 1980). The growing awareness of the risks of STD's has stopped the increase in incidence and a downward trend has been observed in later years. Although we don't have data for our study population on STD's, their incidence appears to be low in most natural fertility settings (Wood, 1994). Next to tubal disease, speculations concern a possible increase of male infertility due to environmental factors. However, the literature on this subject is conflicting. Some studies have found a systematic decline in semen quality (Carlsen *et al.*, 1992) (Auger *et al.*, 1995) (Van Waeleghem *et al.*, 1996) (Irvine *et al.*, 1996), where others found no such decline (Bujan *et al.*, 1996) (Vierula *et al.*, 1996) (Wittmaack and Shapiro, 1992) (Fisch *et al.*, 1996) (Paulsen *et al.*, 1996). Two recent studies (Akre *et al.*, 1999) (Joffe, 2000) that have investigated time trends in overall infertility found that infertility *decreased* in the second half of the 20th century and they came to the same conclusion: if a decline in male fertility is present, it has been more than compensated for by others factors (see also: *chapter 4*).

Finally, couples nowadays probably have more knowledge than in earlier times on the fertile period of the woman and know better how to make maximal use of it, in order to become pregnant. In historical times this knowledge –if at all present- might have been used by women to achieve the opposite: avoid pregnancy. A recent study showed that indeed high pregnancy rates may be achieved by having well-timed intercourse (Gnoth *et al.*, 2003).

In summary, we conclude that childlessness observed in historical 'so-called' natural fertility populations is more frequent than involuntary childlessness in modern times. Most likely, the reasons for this may be a combination of a true decrease in infertility over time, voluntary childlessness in historical populations and selective emigration.

Two estimates for the age-dependency of sterility

We obtained estimates of the age-specific probability of being effectively infertile in two different ways. In *chapter 2 Figure 1*, it was derived from the woman's age at last childbirth, while in *chapter 3 Figure 2*, it was based on the woman's age at marriage, using Cox-regression for the chance to remain childless after marriage. The two curves are shown together in **Figure 3**. The results of the two methods agree for ages until 30. However, the shape of the curves is different: the curve based on age at marriage (*chapter 3*) remains 'flat' until older ages than the curve based on age at last childbirth

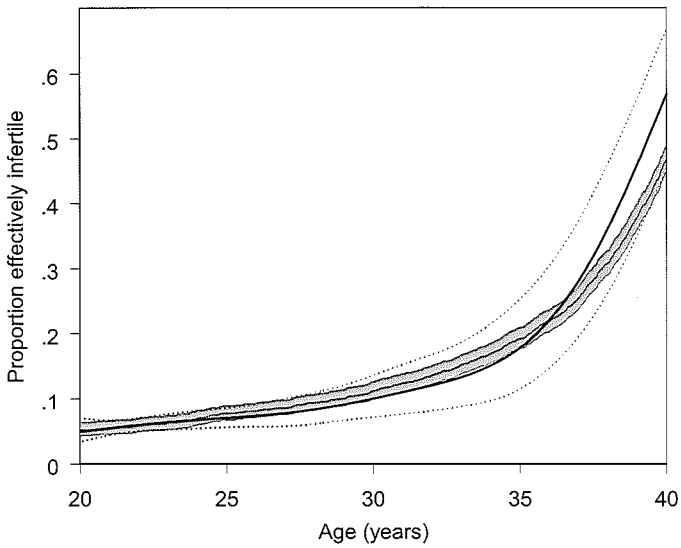


Figure 3: Two estimates (with 95% Confidence Interval) of the age-specific chance to be effectively infertile. (1) Kaplan-Meier curve for the age at last live birth (thin line with shaded area) and (2) using Cox-regression on the time between marriage and birth of the first child (thick line).

(*chapter 2*), while from about age 35 onwards it becomes more steep. Care should be taken not to over-interpret this difference: the uncertainty in the estimates based on the age at marriage (*chapter 3, Figure 2*) is huge for ages above 30, due to the fact that there were only few women who married at these ages (see also *chapter 1*, *Figure 1*). In contrast, the results based on age at last childbirth from *chapter 2* have many women contributing to the estimates at these ages, as is reflected in the narrow 95% confidence band around the curve.

Despite the small statistical uncertainty, some reflection is in place with respect to the results based on age at last childbirth (*chapter 2*). Breastfeeding causes postpartum amenorrhoea, which made the women practically infertile for at least a few months after the birth of the last child. Without breastfeeding, some women might have been able to conceive yet another child, so that the actual age at which they were still able to have a child would be later than their observed age at last childbirth. On the other hand, some woman might already have been effectively infertile while carrying the last child, i.e. before the age at last childbirth. Although these two effects will (at least partly) cancel out, there clearly is extra uncertainty in the curve that is difficult to assess.

Demographers have recognized these problems long ago (Barrett, 1971) (Leridon, 1977). Attempts have been made to obtain the 'true' curve of age at sterility by using computer simulations of the whole process of becoming pregnant and having a live birth. This process is assumed to consist of the age-dependent chance to be definitely

sterile, temporal infertility due to postpartum amenorrhoea, the age-dependent monthly chance to conceive and the age-dependent chance of pregnancy loss (due to intra-uterine death of the foetus) (Trussell and Wilson, 1985). Age at definitive sterility in their analysis is therefore the age at which the woman is no longer able to become pregnant, where in our analysis only pregnancies leading to live birth are considered. While the demographic analysis is probably more precise and may be of value in a theoretical respect, we consider our analysis more meaningful for modern reproductive medicine: it is now widely agreed upon that patients are interested in having a (healthy) child and not just in becoming pregnant.

HISTORICAL DEMOGRAPHIC DATABASES AS A RESEARCH OBJECT FOR CLINICAL REPRODUCTIVE MEDICINE: OPPORTUNITIES AND LIMITATIONS

Observations on natural fertility: a unique opportunity for reproductive medicine

Women in a natural population express two unique aspects of fertility that are not observable in a modern population: 1) the absence of contraception and treatment for infertility gives an undisturbed observation of time to pregnancy leading to live birth and 2) the age at last childbirth marks the end of fertility for a woman. Both aspects are of interest for reproductive medicine, but there are limitations. Biology and behavior are indistinguishable in historical data, and we assume that women were fully expressing their biological capacity. There is good evidence for this: The demographic transition in Quebec did not occur until 1930 and the only method of intentional birth control that may have been practiced was prolonged breastfeeding of children. In this respect, this population is special since it had a tradition of short lactation periods (Lalou, 1990). However, for more detailed analyses of the process of getting children, it data on aspects of fertility such as the frequency of miscarriage, coital frequency and cycle regularity of women would have been of value, but is not available from church registers.

Statistical aspects of analysis on demographic data

One of the consequences of the high fertility of this population is that there were large families. Many women in our selection of subjects had therefore sisters who were also selected, or brothers who were married to other selected women. Therefore, the women in our sample may not be regarded as completely independent, there is clustering within a family. We dealt with this by performing additional analyses, particularly multilevel modeling, in which the level of daughters is nested within the level of mothers.

As mentioned in the introduction (*chapter 1*), the nature of church registers is such that there are inevitably missing data, due to couples moving away (emigration),

missing birth dates of women and missing birth dates of children. For the former two we assumed that they were not related to fertility, while for the latter we had to correct our analysis, to avoid bias in the estimates of childlessness by age.

Topics for further research

The material at hand is well suited for further research on questions from reproductive medicine. Several topics have been suggested during the present research.

1) The role of breastfeeding in the relationship between fertility at young age and age at last childbirth (*chapter* 2) might be investigated further, by using the data of subsequent birth intervals from the same woman. For this aim we might use frailty models, along the way set out by previous work of Vaupel and Larsen (Vaupel, 1988) (Larsen and Vaupel, 1993) (Larsen and Yan, 2000)

2) The fact that we have data on the birth cohorts from 1844 to 1899 implies that several generations are present. The women in our dataset may also have their mother or their daughters in the data. This offers an opportunity to study heritability of parameters of fertility. Technically, multilevel modeling may be used to decompose the variability in fertility of women in several generations.

3) There are several other research questions from reproductive medicine that could be analysed on these data. We mention two of these here.

- Is maternal age related to the fertility of the male offspring? Theories on mitochondrial inheritance predict that sons born of older mother would be less fertile.
- Is paternal age related to the sex-ratio of offspring? (Jacobsen *et al.*, 1999).

Further, questions of a more methodological nature may be further pursued.

4) The observation of the high maternal mortality (*chapter 2*) raises the question of selection of healthy women in the group that stayed married (thus also alive) at least until age 50. Although we concluded in *chapter 2* that the selection was probably not working on fertility, further research is required, e.g. on the parity dependence of maternal mortality.

5) Informative censoring: in *chapter 3* we used the Kaplan-Meier method on time to first childbirths, to estimate the chance of remaining childless by age at marriage. Women whose marriage ended before they had their first child (mainly through death) were considered censored in that analysis. In this chapter we showed that these estimates are higher than simple counts of childlessness in the subgroup of women who remained alive (and married) until age 50. Apparently the censoring of the women who died before age 50 in Kaplan-Meier was informative. Further

investigation in the reasons why women died before age 50 may reveal to what extent the Kaplan-Meier estimates are statistically valid.

CONCLUSIONS AND RECOMMENDATIONS

- The end of female fertility shows huge variation: The chance that a woman is no longer able to conceive a pregnancy leading to live birth at ages 20, 25, 30, 35 and 40 is 5%, 7%, 10%, 17% and 40% respectively.

- Fertility at young age and the age at which fertility ends are positively related, supporting the concept of ovarian aging.

- After 12-24 months of infertility, under 35 years of age, prospects to conceive naturally are on average still good. If diagnostics are started at these durations of infertility (earlier only on indication) and nothing is found, prognosis is better than the population average at that duration of infertility, and therapy could be withheld for the next 1 or 2 years.

References

Akre O, Cnattingius S, Bergstrom R, Kvist U, Trichopoulos D, Ekbom A. (1999) Human fertility does not decline: evidence from Sweden. *Fertil Steril.* 71, 1066-9.

Auger J, Kunstmann JM, Czyglik F, Jouannet P. (1995) Decline in semen quality among fertile men in Paris during the past 20 years. *N Engl J Med.* 332, 281-5.

Barrett JC. (1971) Use of a fertility simulation model to refine measurement techniques. *Demography.* 8, 481-490.

Barrett JC. (1986) The estimation of natural sterility. *Genus.* 42, 23-31.

Bujan L, Mansat A, Pontonnier F, Mieusset R. (1996) Time series analysis of sperm concentration in fertile men in Toulouse, France between 1977 and 1992. *Bmj.* 312, 471-2.

Carlsen E, Giwercman A, Keiding N, Skakkebaek NE. (1992) Evidence for decreasing quality of semen during past 50 years. *Bmj.* 305, 609-13.

Curran JW. (1980) Economic consequences of pelvic inflammatory disease in the United States. *Am J Obstet Gynecol.* 138, 848-51.

Dao TN, Mathieu J, Bouchard JP, De Braekeleer M. (1992) Fertility in myotonic dystrophy in Saguenay-Lac-St-Jean: a historical perspective. *Clin Genet.* 42, 234-9.

De Braekeleer M, Lamarre V, Scriver CR, Larochelle J, Bouchard G. (1990) Fertility in couples heterozygous for the tyrosinemia gene in Saguenay Lac-St-Jean. *Genet Couns.* 1, 259-64.

De Braekeleer M. (1993) A prevalence and fertility study of haemochromatosis in Saguenay-Lac- Saint-Jean. *Ann Hum Biol.* 20, 501-5.

De Braekeleer M, Gauthier S. (1996) Autosomal recessive disorders in Saguenay-Lac-Saint-Jean (Quebec, Canada): a study of inbreeding. *Ann Hum Genet.* 60, 51-6.

Fisch H, Goluboff ET, Olson JH, Feldshuh J, Broder SJ, Barad DH. (1996) Semen analyses in 1,283 men from the United States over a 25-year period: no decline in quality. *Fertil Steril.* 65, 1009-14.

Gnoth C, Godehardt D, Godehardt E, Frank-Herrmann P, Freundl G. (2003) Time to pregnancy: results of the German prospective study and impact on the management of infertility. *Hum Reprod.*, 18, 1959-66.

Greenhall E, Vessey M. (1990) The prevalence of subfertility: a review of the current confusion and a report of two new studies. *Fertil Steril.* 54, 978-83.

Henry L. French statistical work in natural fertility. In: Sheps M, Ridley JC, editors. *Public health and population change.* Pittsburg: University of Pittsburg Press, 1965:333-350.

Irvine S, Cawood E, Richardson D, MacDonald E, Aitken J. (1996) Evidence of deteriorating semen quality in the United Kingdom: birth cohort study in 577 men in Scotland over 11 years. *Bmj.* 312, 467-71.

Jacobsen R, Moller H, Mouritsen A. (1999) Natural variation in the human sex ratio. *Hum Reprod.* 14, 3120-5.

Joffe M. (2000) Time trends in biological fertility in Britain. *Lancet.* 355, 1961-5.

Jones RB, Ardery BR, Hui SL, Cleary RE. (1982) Correlation between serum antichlamydial antibodies and tubal factor as a cause of infertility. *Fertil Steril.* 38, 553-8.

Kane JL, Woodland RM, Forsey T, Darougar S, Elder MG. (1984) Evidence of chlamydial infection in infertile women with and without fallopian tube obstruction. *Fertil Steril.* 42, 843-8.

Lalou R. La haute fécondité des Québécoises avant la révolution contraceptive: effet de nature ou effet de culture? Un bilan des connaissances. Chicoutimi: Université du Québec à Chicoutimi, 1990.

Larsen U, Vaupel JW. (1993) Hutterite fecundability by age and parity: strategies for frailty modeling of event histories. *Demography.* 30, 81-102.

Larsen U, Yan S. (2000) The age pattern of fecundability: an analysis of French Canadian and Hutterite birth histories. *Soc Biol.* 47, 34-50.

Leridon H. (1977) Sur l'estimation de la stérilité. *Population.* 32, 231-48.

Menken J, Trussell J, Larsen U. (1986) Age and infertility. *Science.* 233, 1389-94.

Paulsen CA, Berman NG, Wang C. (1996) Data from men in greater Seattle area reveals no downward trend in semen quality: further evidence that deterioration of semen quality is not geographically uniform. *Fertil Steril.* 65, 1015-20.

Roy R, Declos M, Bouchard G, Mathieu J. (1989) [The reproduction of families affected by Steinert's dystrophy in Saguenay (Quebec), 1885-1971: demographic parameters] La reproduction des familles touchees par la dystrophie de Steinert au Saguenay (Quebec), 1885-1971: parametres demographiques. *Genus.* 45, 65-82.

Tietze C. (1957) Reproductive span and rate of reproduction among hutterite women. *Fertil Steril.* 8, 89-97.

Trussell J, Wilson C. (1985) Sterility in a population with natural fertility. *Population studies.* 39, 269-286.

Van Waeleghem K, De Clercq N, Vermeulen L, Schoonjans F, Comhaire F. (1996) Deterioration of sperm quality in young healthy Belgian men. *Hum Reprod.* 11, 325-9.

Vaupel JW. (1988) Inherited frailty and longevity. *Demography.* 25, 277-87.

Veillette S, Perron M, Mathieu J. (1989) [Myotonic dystrophy: II. Marital status, fertility and gene transmission] La dystrophie myotonique: II. Nuptialite, fecondite et transmission. *Can J Neurol Sci.* 16, 114-8.

Vierula M, Niemi M, Keiski A, Saaranen M, Saarikoski S, Suominen J. (1996) High and unchanged sperm counts of Finnish men. *Int J Androl.* 19, 11-7.

Vincent P. (1950) La stérilité physiologique des populations. *Population.* 5, 45-64.

Wittmaack FM, Shapiro SS. (1992) Longitudinal study of semen quality in Wisconsin men over one decade. *Wis Med J.* 91, 477-9.

Wood JW. (1994) *Dynamics of human reproduction: biology, biometry, demography*, Aldine de Gruyter,New York.

Discussion and conclusions part II

Part II of this thesis addresses research questions related to the development of an individualized treatment protocol in patients with WHO II anovulatory infertility. In this chapter, we summarize the answers to these questions and discuss the implications for the treatment protocol.

ANSWERS TO THE RESEARCH QUESTIONS

1. **Which patient characteristics are predictive of pregnancy chances following ovulation induction with Clomiphene Citrate (CC)? What is -in this respect- the relation between the intermediate endpoint 'ovulation' and the final endpoint pregnancy?**

Three analyses were performed: 1) prediction of the chance to remain anovulatory (i.e. to be CRA 2) prediction of pregnancy, given that the patient was not CRA and 3) integration of 1) and 2).

For the prediction of the chance to be CRA (1), a multivariable logistic regression model was built, including the Free Androgen Index (FAI), Body Mass Index (BMI), cycle history, and mean ovarian volume as predictors. Patients whose ovaries are less likely to respond to stimulation by FSH due to CC treatment had elevated FAI, higher BMI, amenorrhea instead of oligomenorrhea and large mean ovarian volume. The prediction model for the pregnancy chances in patients who were not CRA (2) was a Cox regression model and contained two predictors: patients who became pregnant presented more frequently with lower age and amenorrhea upon initial screening. The two models (1) and (2) were combined in a graphical tool (3) to facilitate the calculation of the probability of pregnancy leading to live birth with CC treatment. The opposite effects of amenorrhea versus oligomenorrhea (a negative effect in (1), a positive effect in (2)) more or less cancelled out.

2. **Which patient characteristics are predictive of pregnancy chances of classical ovulation induction with Clomiphene Citrate (CC) followed by Gonadotrophin ovulation induction (GOI)?**

A Cox regression model was built containing the woman's age, the insulin/glucose ratio and the duration of infertility as predictors of pregnancy leading to singleton live birth. The effects of factors that were important in predicting the outcome of CC treatment alone, such as FAI and BMI, apparently disappear with the use of GOI after CC.

3. **What is the optimal treatment sequence for WHO II patients based on their individual characteristics and on the response in previous treatment steps?**

A cost-effectiveness analysis was performed comparing several alternative treatment strategies with the reference strategy CC+GOI+IVF, for 16 different patient groups. The alternative strategies were GOI + IVF (i.e. skip CC), CC + IVF (i.e. skip GOI), CC (i.e. stop after unsuccessful CC), CC + GOI (stop after unsuccessful GOI) and finally no treatment at all. In the latter three strategies it was assumed that patients would try to become pregnant spontaneously. The maximal acceptable threshold for the extra costs per extra ongoing pregnancy was assumed to be € 10.000.

For most patients, the reference strategy CC+GOI+IVF is cost-effective. An exception may be made for women over 30 years of age who have elevated androgen levels. For them, the treatment strategy CC + IVF (i.e. skip GOI) is indicated, as the addition of GOI produces only few extra pregnancies against costs of over 90,000 Euro per pregnancy. The reference strategy is therefore not acceptable for this patient group, which comprises about 14% of the group of patients that is not pregnant after CC and who would start GOI treatment. The strong association between outcome of treatment and obesity (for CC from our own analysis, for gonadotrophins mainly from literature) implies that chances could be improved by reducing overweight. This concerns 30-40% of the WHO II group.

METHODOLOGICAL CONSIDERATIONS

A few remarks should be made in retrospect with regards to methodological issues. They concern both the general aspects of prognostic model validation, as well as the more specific issue of the role of the intermediate outcome CRA in the analysis of time to pregnancy.

Prognostic model validation

In *chapter 6* we formulated recommendations for the development of prognostic models, including internal validation and shrinkage of regression coefficients as a final step. We followed these recommendations in *chapter 7*, when we developed the prediction model for CRA. However, we did not perform an internal validation of the

Cox-regression model for pregnancy chances in non-CRA patients in *chapter 8*. At the time of development of that model, no standard statistical package was accessible to perform these procedures for Cox-models. Only later on, when the model was used in *chapter 9* to be combined with the CRA-model from *chapter 7*, we established a shrinkage factor by performing a bootstrap procedure. The details were omitted from *chapter 9* and will be given here. The model was built using backward stepwise selection of variables starting with 7 candidate predictors. In the final model only 2 of these were retained, *age* and *type of cycle disturbance (oligo- or amenorrhea)*. The bootstrap procedure (with 200 replications) revealed that the shrinkage factor, necessary to produce well-calibrated predictions in future patients, was 0.84. Further, the apparent discriminative ability of the model, measured by a c-statistic of 0.68, was corrected for optimism by the bootstrap procedure to a value of 0.65. The shrinkage factor was applied in the development of the nomogram in *chapter 9*, as well as in the cost-effectiveness analysis in *chapter 11*.

So far, none of the prediction models in this thesis has been externally validated. Nevertheless, we used the models in a cost-effectiveness analysis to formulate recommendations for the treatment protocol of WHO II anovulatory infertile patients (*chapter 11*). Because only internal validity has been assessed we may only rely on the results of this cost-effectiveness analysis for patients that are similar and from a similar setting. Future studies should be designed to perform external validation of the prediction models in independent patient groups.

Analysis of time to pregnancy

As stated in *chapter 6*, the analysis for research question 1 -the prediction of time to pregnancy following CC treatment- was split into three separate analyses: one for the probability to be CRA (*chapter 7*), one for time to pregnancy given that the patient is not CRA (*chapter 8*), and finally the integration of these two (*chapter 97*). If we had performed a Cox regression analysis for time to pregnancy on the whole group of CC patients, the CRA patients would become censored from the moment they are identified as such. The reason for this is that CRA is an intermediate outcome that leads to stopping CC-treatment, because a patient who is CRA has no chance of pregnancy with this medication. However, this implies that the censoring would be informative, violating a basic assumption of Cox regression. Cox regression assumes that patients who are censored at a certain follow-up time have the same chance to get the event of interest (become pregnant) after that time, as do the patients who are not censored. The result of such an analysis would be that the cumulative probability to be pregnant would be overestimated.

In *chapter 10*, we used Cox regression to predict the chance of pregnancy from the start of CC treatment onwards and continuing until the treatment with GOI was stopped. In contrast to the analysis for the prediction of pregnancy with CC, which was done in two steps, we did this analysis in one step. Patients who after three unsuccessful cycles of CC-treatment are identified as CRA, from that moment on do

not have a zero chance to become pregnant, since they will start treatment with GOI, which is included in this analysis. Therefore, informative censoring is not an issue for CRA patients in this analysis. One could argue that the group at risk in the Cox regression from the moment of identifying patients who are CRA (i.e. after 3 cycles of CC treatment) will consist of two groups of patients: patients who are non-CRA and still continuing CC, together with the CRA patients who have started GOI treatment: their pregnancy chances may be different. However, this non-homogeneity in chances is different from informative censoring: No assumption of the analysis is violated here and the estimate will give a proper account of the experience of the group of patients that started treatment.

CONCLUSIONS AND RECOMMENDATIONS

- For most patients, the current standard treatment protocol -start with CC and proceed with GOI followed by IVF, in case no pregnancy has been achieved- is cost effective.

- An exception may be made for women above 30 years of age with elevated androgen levels, for whom the treatment with gonadotrophins could be skipped: performing GOI in these women gives only a small increase in pregnancy chances against costs of over 90,000 Euro per extra pregnancy.

- The strong association between outcome of treatment and obesity (for CC from our own analysis, for gonadotrophins mainly from literature) implies that chances could be improved by reducing overweight. This concerns 30-40% of the WHO II group.

Summary

This thesis is devoted to chances in fertility. It treats two subjects that are of interest to clinical reproductive medicine: fertility chances in a historical population and chances of success of treatment in one particular group of patients: those with anovulatory infertility.

Part I of the thesis describes studies in a 19th century natural fertility population on variation between individuals in fertility chances. Two research questions are investigated with the aim to learn something from the past: one concerns the variation in the age at which female fertility comes to an end, and seeks evidence whether this variation is consistent with the concept of an underlying biological process of ovarian ageing. The second question concerns the pregnancy chances after a period of unsuccessful trying. This is of interest to clinicians who have to decide to start treatment for a given patient.

Chapter 1 introduces the 19th century population of the Saguenay Lac St.-Jean region in Quebec Canada. The historical and social/religious (Roman-Catholic) context of this region are briefly reviewed, the concept of natural fertility is explained and a demographic analysis of the data is presented that shows that the study population may be considered a natural fertility population. We selected 4,125 married women from this population, who were born in the region between 1844 and 1900, and who did not emigrate.

In **Chapter 2**, the first research question is analyzed: variability in female reproductive life span is evaluated and the association between the reproductive life span and fertility at young age is studied. To be able to assess the reproductive life span, a woman should be followed at least until an age at which it is certain that she will not have any more children. We therefore selected the 2,488 women whose marriage lasted at least until age 50. The chance that a woman no longer bares a child was 10% at age 29 years. It increased to 25%, 40%, 50%, 60% and 75% at ages 37, 39, 40.5, 41 and 42.5 respectively, and was already 90% at age 44 years. There was a significant correlation ($r = 0.15$, $P < 0.001$) between fertility at young age and age at last childbirth: the more children a woman had before her 30th birthday, the longer she

was able to go on having children after age 30. This finding is consistent with the theory of ovarian ageing.

The subject of **Chapter 3** is inspired by the clinical question of when to start treatment in a couple with primary or secondary infertility. To answer this question, a reliable estimate of the spontaneous pregnancy chances is needed, taking into account the woman's age, the duration of infertility and parity. To obtain these estimates, we followed the women in our study population from marriage onwards. Women who had not conceived their first child within 1 to 4 years after marriage or their second child within 1 to 4 years after the first childbirth were then labeled as primary or secondary infertile. Of the women with 'primary' infertility of one year duration, 44% were found to conceive a child within the next year in the age group 20-24, and 39%, 32% and 22% respectively in the age groups 25-29, 30-34 and 35-39. At 2 years durations the age specific predictions were 26%, 21%, 17% and 12%. Predictions for 'secondary' infertility at one year duration were 63%, 56%, 48% and 36%, and at 2 years duration 38%, 33%, 28% and 19% respectively in the age groups 20-24, 25-29, 30-34 and 35-39. We concluded that in women under 35, after one year of primary infertility and 1-2 years of secondary infertility, prospects are on average fairly good. Translated to modern times, we could say that if the diagnostic work-up of the couple does not point to adverse factors influencing pregnancy chances, their prognosis may be better than the average estimate that we found here, and therapy could be withheld for the next 1 or 2 years.

The final chapter of Part I, **Chapter 4**, contains a short exposé on the concept of time to pregnancy and how this concept relates to monthly pregnancy chances. Based on studies using the concept of time to pregnancy, the conclusion is drawn that the near panic sometimes expressed in the lay press about the effects of environmental pollution on sperm quality and male fertility is not justified.

The subject of Part II is the prediction of pregnancy chances after treatment of patients who have fertility problems due to cycle disturbances (normogonadotropic anovulatory infertility, classified as WHO group 2). The current treatment protocol for these patients consists in the first line of ovulation induction with Clomiphene Citrate (CC), a simple and cheap medicine. When the patient doesn't become pregnant with this medication, the second line treatment, gonadotrophin ovulation induction (GOI) is started. This treatment is more burdensome, and also more costly than CC. The final treatment is In Vitro Fertilization (IVF). Patients differ in their response to each of these treatment modalities. The aim here is to construct tools that may be used to predict -for individual patients- the pregnancy chances with the standard treatment protocol. These prediction tools are then used in an analysis in which the efficiency of the current treatment protocol is investigated and compared to several alternatives. The background and aims of this part of the thesis are briefly introduced in **Chapter 5**. All subsequent chapters are based on the prospectively collected data of a consecutive

series of these patients in ErasmusMC Rotterdam, included between 1992 and 1999, with ongoing follow-up.

Chapter 6 presents a tutorial on the methodology of prognostic modeling in fertility. Prognostic evidence is summarized in a statistical model that relates individual characteristics with the possible outcomes. Such a model will provide estimates of prognosis for individualized patient profiles, allowing new patients to profit from the experience of the cohort of previous patients used to build the model. We present and discuss characteristics of good prognostic evidence for clinical use, focusing on study design, statistical analysis, evaluation, and presentation of results. As an example, this chapter discusses the prediction of time to pregnancy following induction of ovulation with CC. This prediction was broken down in two steps for methodological reasons, leading to two separate prognostic models. The first model predicts an intermediate outcome, the chance that the patient will have Clomiphene Resistant Anovulation (CRA), i.e., no ovulation in response to CC medication; the second model predicts the final outcome (time until pregnancy) in women who do ovulate.

The next three chapters are the elaboration of the ideas expressed in *Chapter 6*, and concern the prediction of outcome after ovulation induction with Clomiphene Citrate (CC), based on patient characteristics assessed during standardized initial screening. To begin with, in **Chapter 7**, a prediction model of the chance that a patient will remain anovulatory during CC treatment (CRA) is developed. This study used the data of 201 patients presenting with oligomenorrhea or amenorrhea and infertility. After a complete follow-up (in the case of a no response, at least 3 treatment cycles with daily CC doses up to 150 mg), 156 patients (78%) ovulated. Patients whose ovaries are less likely to respond to CC treatment can be predicted on the basis of initial screening characteristics, such as Free Androgens Index (FAI: testosterone : 'sex hormone-binding globulin' ratio), the body mass index (BMI), cycle history (oligomenorrhea or amenorrhea), and mean ovarian volume. The multivariable prediction model based on these four factors had an 'area under the receiver operating characteristics curve' of 0.82.

Next, in **Chapter 8**, the prediction of the step from ovulation to pregnancy was undertaken. A multivariable prediction model was built on the data of 160 women who reached ovulation after Clomiphene Citrate (CC) medication. This was a slight extension of the group of patients who reached ovulation during CC treatment in *Chapter 7*. A cumulative conception rate of 73% was reached within 9 CC-induced ovulatory cycles. Age and cycle history (oligomenorrhea vs. amenorrhea) were identified as the most important parameters for prediction of conception. Once they ovulate, there is more to be gained from CC ovulation induction by younger women presenting with amenorrhea. We see here that having amenorrhea has a positive effect on pregnancy chances in ovulatory women, whereas we have seen in *Chapter 7* that amenorrhea is a negative risk factor to reach ovulation.

The objective of the last chapter in this trilogy on outcome after CC, **Chapter 9**, is to establish whether initial screening characteristics predict live birth after Clomiphene Citrate (CC) induction of ovulation. To this end, the prediction model for the chance to be CRA from *Chapter 7* was combined with an adapted version of the prediction model for pregnancy chances in ovulatory women from *Chapter 8*. Based on this integrated analysis, the cumulative live birth rate within 12 months was 42% for the total study population. The combined prediction model contains four variables: free androgen index (testosterone sex hormone-binding globulin ratio), body mass index, cycle history (oligomenorrhea versus amenorrhea) and woman's age. For convenient use of the model in practice, it was represented in the form of a nomogram. Applying this nomogram in the clinic may be a step forward in optimizing the decision-making process in the treatment of normogonadotropic anovulatory infertility. Alternative first line treatment options could be considered for some women with limited chances for success.

The first two steps of the current protocol may be called classical ovulation induction. Because the overall chances of success following classical ovulation induction are not well established, a shift in first line therapy can be observed towards alternative treatments. In **Chapter 10**, our aim was to (1) reliably assess the probability of singleton live birth following classical induction of ovulation and (2) construct a prediction model, based on individual patient characteristics assessed upon standardized initial screening, to help identify patients with poor chances of success with classical ovulation induction. The cumulative singleton pregnancy rate (leading to live birth) after 12 and 24 months of follow-up was 50% and 71%, respectively. A multivariable Cox regression model was constructed, containing the woman's age, the insulin/glucose ratio and the duration of infertility. With a cut-off value of 30% for low chance, the model predicted probabilities at 12 months lower than this cut-off for 25 out of 240 (10%) patients. We conclude that classical ovulation induction gives very good results in normogonadotrophic anovulatory infertility. Alternative treatment options are not indicated as first line therapy in these patients, except for subgroups with poor prognosis. These women can be identified by older age, longer duration of infertility and higher insulin/glucose ratio.

The standard treatment algorithm in normogonadotrophic anovulatory infertility consists of Clomiphene citrate (CC) as least burdensome and cheapest first line, exogenous gonadotrophin ovulation induction (GOI) as second line and In Vitro Fertilization (IVF) as the most costly and most burdensome last resort. Response to these treatment modalities is highly variable, but may be predicted on the basis of individual patient characteristics, as we have seen in our studies. In **Chapter 11**, our aim was to devise a patient-tailored and cost-effective treatment algorithm, based on individual patient characteristics. Sixteen different patient subgroups are defined, according to the presence or absence of the following 4 risk factors: age > 30, amenorrhea, elevated androgens and obesity. For each subgroup, the chances of response with each of the three treatments, and their costs, were calculated using the

previously developed prediction models. The costs per pregnancy of the reference strategy CC+GOI+IVF were compared with three alternative strategies: one in which the CC step is skipped (i.e. start with GOI and after that go to IVF), one without the GOI-step (i.e. start with CC and after that go directly to IVF), and the final one in which no treatment is started. For most patients, the current standard treatment protocol is cost-effective, at a threshold-value for the cost-effectiveness ratio of € 10.000 per extra ongoing pregnancy. An exception should be made for women above 30 years of age with elevated androgen levels, for whom the treatment with gonadotrophins could be skipped. Performing GOI in these women gives only a small increase in pregnancy chances against costs of over 90,000 Euro per extra pregnancy. Finally, the strong association between outcome of treatment and obesity (for CC from our own analysis, for gonadotrophins mainly from literature) implies that chances could be improved by reducing overweight. This concerns 30-40% of the WHO 2 group.

In conclusion we have demonstrated in Part I that the end of female fertility is very variable and that fertility at young age and the age at which fertility ends are positively related, consistent with the concept of ovarian ageing. Next, we have found that for women under 35 years of age, after 12-24 months of infertility, prospects to conceive naturally are on average still good. This could give support to a conservative attitude towards early treatment.

In Part II we have seen that for most patients, the current standard treatment protocol -start with CC and proceed with GOI followed by IVF, in case no pregnancy has been achieved- is cost effective. An exception may be made for women above 30 years of age with elevated androgen levels, for whom the treatment with gonadotrophins could be skipped. The results are critically discussed in **Chapter 12** and recommendations for further research are given there.

Samenvatting

Dit proefschrift gaat over kansen op het gebied van vruchtbaarheid. Er worden twee onderwerpen behandeld die van belang zijn voor de klinische praktijk van de voortplantingsgeneeskunde: vruchtbaarheidskansen in een historische populatie en kansen op succes van behandeling bij één bepaalde groep van infertiliteitspatiënten, namelijk die met ovulatiestoornissen.

Deel I van het proefschrift beschrijft studies in een 19-de eeuwse natuurlijke populatie over vruchtbaarheidskansen en de variatie daarin tussen individuen. Er worden twee onderzoekvragen uitgewerkt, met als doel een les uit het verleden te leren: de eerste vraag gaat over de variatie in de leeftijd waarop de vruchtbaarheid van de vrouw eindigt en zoekt naar een mogelijke verklaring van die variatie die verband houdt met het biologisch proces van ovariële veroudering. De tweede vraag gaat over de zwangerschapskansen na een bepaalde periode van onvervulde kinderwens. Deze vraag is van belang voor klinici die moeten beslissen of zij al dan niet gaan behandelen bij een gegeven patiënt.

In **Hoofdstuk 1** wordt de 19-de eeuwse populatie uit het Saguenay Lac St. Jean gebied in Quebec, Canada geïntroduceerd. We bespreken de historische en sociaal-religieuze (Rooms-Katholieke) achtergrond van dit gebied, geven een definitie van wat het concept natuurlijke vruchtbaarheid inhoudt en presenteren de resultaten van een demografische analyse die laat zien dat deze populatie als een natuurlijke vruchtbaarheidspopulatie mag worden beschouwd. Voor ons onderzoek werden uit deze populatie 4.125 getrouwde vrouwen geselecteerd, die in het gebied zelf waren geboren tussen 1844 and 1900, en die niet emigreerden.

In **Hoofdstuk 2** wordt de eerste onderzoeksvraag behandeld. De variatie in 'reproductieve levensduur' van de vrouw is onderzocht en de samenhang met de vruchtbaarheid op jonge leeftijd werd bestudeerd. Met 'reproductieve levensduur' wordt bedoeld tot op welke leeftijd een vrouw nog in staat is om kinderen te krijgen. Om dit goed te kunnen vaststellen dienen de vrouwen lang genoeg in leven, én getrouwd te blijven. Daarom werd de selectie uit de populatie beperkt tot de 2.488 vrouwen die nog in leven én nog getrouwd waren op hun 50-ste verjaardag. Zo'n 10%

van de vrouwen kreeg geen kinderen meer na haar 29-ste. Dit percentage lag op 25%, 50% en 75% op de leeftijden 37, 40,5 en 42,5 respectievelijk, en was al 90% voor vrouwen van 44. De leeftijd waarop de vrouwen hun laatste kind kregen hing samen met een maat voor hun vruchtbaarheid op jonge leeftijd ($r = 0.15$, $P < 0.001$): hoe meer kinderen een vrouw voor haar 30-ste kreeg, hoe langer ze na haar 30-ste door kon gaan met kinderen krijgen. Deze bevinding is consistent met de theorie van ovariële veroudering.

Hoofdstuk 3 heeft als doel vast te stellen hoe hoog de spontane zwangerschapskansen zijn van paren met onvruchtbaarheid, en hoe deze kansen samenhangen met de leeftijd van de vrouw, de duur van de onvruchtbaarheid en of het paar al dan niet kinderen heeft. Hiertoe werden de vrouwen gevolgd vanaf het huwelijk. Vrouwen die binnen 1 tot 4 jaren na het huwelijk nog niet zwanger waren van hun eerste kind, danwel van hun tweede kind binnen 1 tot 4 jaren na de geboorte van hun eerste kind, werden als primair respectievelijk secundair infertiel beschouwd, volgens de klinische definitie die tegenwoordig wordt gehanteerd. In de leeftijdsgroep 20 tot 24 jaar bleek 44% van de vrouwen met primaire infertiliteit met een duur van 1 jaar in het daarop volgende jaar zwanger te worden. Dit percentage lag op 39%, 32% en 22% respectievelijk in de leeftijdgroepen 25-29, 30-34 en 35-39. Bij een infertiliteitsduur van 2 jaar waren de leeftijds-specifieke percentages 26%, 21%, 17% en 12%. De kansen voor vrouwen met secundaire infertiliteit lagen een stuk hoger: na 1 jaar secundaire infertiliteit was 63%, 56%, 48% en 36% zwanger in het volgende jaar, en bij een infertiliteitsduur van 2 jaar was dit 38%, 33%, 28% en 19% respectievelijk in de genoemde leeftijdsklassen. Onze conclusie is dat voor vrouwen onder de 35 jaar, bij een infertiliteitsduur van één jaar, en 1-2 jaar bij secundaire infertiliteit, de vooruitzichten gemiddeld genomen nog redelijk goed zijn. Vertaald naar de moderne tijd betekent dit dat als er bij een gegeven paar niets wordt gevonden dat wijst op een slechte prognose, de kansen waarschijnlijk beter zijn dan hier gegeven, en dat overwogen kan worden om behandeling nog 1 tot 2 jaar uit te stellen.

In het laatste hoofdstuk van **Deel I**, **Hoofdstuk 4**, wordt een kort exposé gegeven over het belangrijke concept van tijd tot aan zwangerschap ('time to pregnancy') en hoe dit concept is gerelateerd aan de maandelijkse kans om zwanger te worden. Op basis van gepubliceerde studies die dit concept hebben gehanteerd, wordt geconcludeerd dat de vruchtbaarheid in de algemene bevolking de laatste decennia eerder is toegenomen dan afgenomen. De paniek die soms wordt geuit in de 'lekenpers' over de effecten van milieuvervuiling op de kwaliteit van het sperma en de mannelijke vruchtbaarheid, lijkt dan ook niet terecht.

Het onderwerp van **Deel II** van dit proefschrift is predictie van zwangerschapskansen na behandeling van patiënten met een vruchtbaarheidsprobleem ten gevolge van cyclusstoornissen (normogonadofe anovulatoire infertiliteit, WHO groep 2). Het huidige behandelprotocol voor deze patiënten bestaat uit een sequentie van drie behandelstappen. De eerstelijnsbehandeling is ovulatie-inductie met Clomipheen

Citraat (CC), een weinig belastend en goedkoop medicijn. Als de patiënt hiermee niet zwanger wordt binnen een bepaald aantal maanden, begint men doorgaans met de tweedelijnsbehandeling, Gonadotrofinen Ovulatie Inductie (GOI). Deze behandeling is aanmerkelijk belastender en ook duurder dan CC. De laatste behandelingsoptie is In Vitro Fertilisatie (IVF). De patiënten verschillen nogal in hun respons op elk van deze behandelingen. Het doel is hier om materiaal aan te dragen dat kan helpen om voor individuële patiënten de zwangerschapskansen met dit behandelprotocol in te schatten. Deze predictie-tools worden vervolgens gebruikt in een analyse waarin de efficiëntie van het huidige behandelprotocol wordt vergeleken met verschillende alternatieve protocollen. De achtergrond en doelstelling van dit deel vormen het onderwerp van **Hoofdstuk 5**. Alle verdere hoofdstukken in dit deel zijn gebaseerd op de prospectief verzamelde data van een consecutieve reeks patiënten in ErasmusMC Rotterdam. De inclusie liep van 1992 tot en met 1999, met doorgaande follow-up daarna.

Hoofdstuk 6 is een tutorial over de methodologie van prognostisch modelleren in de fertiliteit. In het algemeen wordt prognostische evidentie samengevat in de vorm van een statistisch model dat de relatie beschrijft tussen individuele patiëntkarakteristieken en de mogelijke uitkomsten. Zo'n model geeft dan schattingen van de prognose voor individuele patiëntprofielen, waarmee nieuwe patiënten kunnen profiteren van de ervaring van de groep patiënten die is gebruikt bij het bouwen van het model. We geven en bediscussiëren kenmerken van goede prognostische evidentie, met name geschikte studie-designs, statistische analyse, evaluatie, en presentatie van de resultaten. Als illustratief voorbeeld wordt de predictie van zwangerschapskansen na ovulatie-inductie met CC gebruikt. Dit predictie-probleem werd opgesplitst in 2 stappen, omwille van methodologische redenen, wat leidde tot twee predictiemodellen. Het eerste model voorspelt een tussengelegen uitkomst, de kans dat de patiënt niet zal reageren met een ovulatie (eisprong) op het medicijn (wat Clomiphene Resistant Anovulation (CRA) wordt genoemd); het tweede model voorspelt de uiteindelijke uitkomst (zwangerschap) voor de vrouwen die wél ovuleren met CC. Tot slot worden deze twee modellen geïntegreerd in één predictiemodel.

De volgende drie hoofdstukken bevatten de concrete uitwerking van de ideeën die in *Hoofdstuk 6* zijn neergezet met betrekking tot de predictie van de uitkomst van ovulatie-inductie met Clomipheen Citraat (CC). Om te beginnen wordt in **Hoofdstuk 7** een predictie-model besproken voor de kans dat een patiënt geen ovulatie bereikt tijdens de behandeling met CC (i.e. de patiënt is CRA). Deze studie gebruikte de gegevens van 201 patiënten met oligomenorroe (cylcusduur > 35 dagen) of amenorroe (geen cyclus, of cyclusduur > 6 maanden) en infertiliteit. Na een volledige follow-up (wanneer er geen ovulatie optreedt: tenminste 3 behandelcycli met dagelijkse CC-dosis oplopend tot 150 mg/dag) waren 156 patiënten (78%) ovulatoir. Patiënten hadden een lagere kans op ovulatie met CC naarmate ze hogere waarden hadden van de factoren vrije androgeen-index (FAI: testosteron : 'sex hormone-binding globulin' ratio), de body mass index (BMI) en gemiddeld volume van de ovaria, en wanneer ze

amenorroe in plaats van oligomenorroe hadden. Het predictiemodel met deze 4 factoren had een 'area under the receiver operating characteristics curve' van 0.82.

Vervolgens werd in **Hoofdstuk 8** een predictiemodel gebouwd voor zwangerschapskansen na ovulatie, op de data van 160 vrouwen die een ovulatie hadden bereikt met CC. Dit was een kleine uitbreiding van de data van de patiënten die ovulatie bereikten met CC uit *Hoofdstuk 7*. De 'cumulative conception rate' na 9 CC-induceerde ovulatoire cycli was 73%. De leeftijd en oligomenorrhea versus amenorrhea waren de belangrijkste voorspellende parameters voor de kans op conceptie. Het bleek dat –als ze eenmaal ovuleerden met CC- jongere vrouwen en vrouwen met amenorroe de hoogste kans op zwangerschap hadden. We zien hier dat het hebben van amenorroe in plaats van oligomenorroe een positief effect heeft op zwangerschapskansen in vrouwen die ovuleren met CC, terwijl in *Hoofdstuk 7* bleek dat het hebben van amenorroe juist een negatieve factor is voor het bereiken van ovulatie.

Het doel van het laatste hoofdstuk in deze trilogie over de uitkomst van CC-behandeling, **Hoofdstuk 9**, is om een predictiemodel op basis van initiële patiëntkarakteristieken te ontwerpen voor de zwangerschapskansen met CC ovulatie-inductie. Het predictiemodel voor de kans om CRA te zijn, uit *Hoofdstuk 7*, werd gecombineerd met een aangepaste versie van het predictiemodel voor zwangerschapskansen in ovulatoire vrouwen uit *Hoofdstuk 8*. Op basis van deze geïntegreerde analyse was de cumulatieve kans op het bereiken van een zwangerschap leidend tot een levendgeboren kind 42% voor de hele groep van patiënten in de studie. Het gecombineerde predictiemodel bevat vier variabelen: FAI, BMI, oligomenorroe versus amenorroe en de leeftijd van de vrouw. Om het model toegankelijk te maken voor gebruik in de praktijk is er een nomogram van gemaakt. Toepassing van dit nomogram zou een stap voorwaarts kunnen zijn tot het optimaliseren van de behandeling van patiënten met WHO 2 anovulatoire infertiliteit.

De eerste twee stappen uit het huidige behandelprotocol vormen samen de klassieke methode van ovulatie-inductie. Omdat nooit goed is vastgesteld wat hiervan de kans op succes is, lijkt er een verschuiving op te treden in eerstelijnstherapie naar andere behandelmethodes. In **Hoofdstuk 10** stellen we ons ten doel om (1) een betrouwbare schatting te verkrijgen van de kans op een eenling zwangerschap (leidend tot een levendgeboren kind) met klassieke ovulatie-inductie en (2) een predictiemodel te maken op basis van individuële patiëntkarakteristieken, om patiënten met een lage kans op succes met deze behandelingen te helpen identificeren. De cumulatieve kans op een eenlingzwangerschap (leidend tot levendgeboren kind) bedroeg 50% en 71%, na 12, respectievelijk 24 maanden follow-up. De statistische analyse leverde een predictiemodel op met als voorspellende factoren de leeftijd van de vrouw, de uit bloed-serum bepaalde insuline/glucose ratio en de duur van de infertiliteit. Wanneer we de kans op zwangerschap met deze behandelingen als laag definiëren wanneer die onder de 30% komt, voorspelt het model lage kansen na 12 maanden voor 25 van de 240 (10%) patiënten. Samengevat concluderen we dat klassieke ovulatie-inductie zeer

goede resultaten geeft in normogonadotrope anovulatoire infertiliteit. Andere behandelingsopties zijn niet geindiceerd voor deze groep patiënten, behalve misschien voor sub-groepen met een slechte prognose. Deze vrouwen kunnen worden gekenschetst door een hogere leeftijd, langere duur van de infertiliteit en een hogere insuline/glucose ratio.

In **Hoofdstuk 11** komen we toe aan het uitwerken van de doelstelling dat met behulp van predictiemodellen de keuze en volgorde van de drie behandelingen misschien efficiënter kan worden gemaakt. De predictiemodellen in de voorgaande hoofdstukken vormen hiervoor de basis. We hebben 16 verschillende prognostische sub-groepen gevormd door het combineren van de 4 belangrijkste voorspellende factoren uit de voorgaande analyses en uit de literatuur, te weten: leeftijd > 30, amenorroe, verhoogde mannelijke-hormoonspiegels en overgewicht. De kansen op zwangerschap met elk van de drie behandelingen en de kosten van behandeling werden berekend voor alle 16 patient-profielen. De kosten per doorgaande zwangerschap van de referentie-strategie CC+GOI+IVF zijn vervolgens vergeleken met drie alternatieve strategieën: één waarin de CC-stap wordt overgeslagen, één met overslaan van de GOI-stap en één waarin wordt afgezien van (verdere) behandeling. Voor de meeste patiënten is het standaard protocol kosten-effectief, wanneer een grenswaarde wordt gehanteerd van € 10.000 per extra doorgaande zwangerschap. Er kan wellicht een uitzondering worden gemaakt voor vrouwen boven de 30 jaar met verhoogde mannelijke-hormoonspiegels, voor wie de GOI-behandeling overgeslagen zou kunnen worden. Tot slot suggereert de sterk negatieve rol die overgewicht speelt in het bepalen van de kans op succes (voor CC in onze eigen analyses, voor GOI vooral in de literatuur) dat gewichtsreductie een relatief goedkope en effectieve methode kan zijn om de kansen op succes van behandeling te vergroten. Dit betreft zo'n 30 - 40% van de WHO 2 groep.

Concluderend kunnen we stellen dat **Deel I** van dit proefschrift laat zien dat in een natuurlijke vruchtbaarheidspopulatie, de leeftijd waarop de vruchtbaarheid van een vrouw eindigt grote variatie vertoont en dat deze leeftijd hoger is naarmate de vrouw op jongere leeftijd een hogere vruchtbaarheid had. Deze bevinding is consistent met het concept van 'ovariële veroudering'. Daarnaast vonden we dat vrouwen jonger dan 35 jaar, die 12 tot 24 maanden zonder succes hebben gepoogd zwanger te worden, nog steeds heel redelijke kansen hebben. Deze bevinding geeft steun aan een terughoudend beleid m.b.t. vroegtijdige behandeling.

In **Deel II** hebben we gezien dat voor de meeste patiënten met WHO 2 anovulatoire infertiliteit, het huidige standaard behandelprotocol –beginnen met Clomifeen Citraat, doorgaan met Gonadotrofine Ovulatie-Inductie (GOI), en tenslotte In Virto Fertilisatie (IVF)– kosten-effectief is. Een uitzondering kan worden gemaakt voor vrouwen van boven de 30 met verhoogde mannelijke-hormoonspiegels. Bij hen zou de GOI behandeling kunnen worden overgeslagen. De resultaten worden kritisch

bediscussieerd in **Hoofdstuk 12** en daar worden ook aanbevelingen voor nader onderzoek gegeven.

Dankwoord

Mijn promotie tot doctor beschouw ik als een late roeping. Na lange tijd met veel plezier aan verschillende projecten te hebben gewerkt op de afdeling Maatschappelijke Gezondheidszorg, zonder doelgericht naar een promotie te streven, heeft het op een of andere manier toch zover moeten komen. Ik geloof meer in de samenloop der dingen dan in voorbestemdheid. Toch waren hier naar mijn idee enkele sturende krachten in het spel, die ik nu zeer erkentelijk ben voor hun bijdrage.

Allereerst en bovenal was dat Dik Habbema, die met veel geduld dit langdurige proces heeft begeleid: als het aan jou had gelegen zou ik waarschijnlijk al drie keer zijn gepromoveerd. Dik, ik waardeer je steun, je scherpzinnigheid en je bereidheid in de laatste fase zo veel tijd voor mij vrij te maken, zelfs in de avonduren. Ik wil daarbij ook Annelies betrekken: dank voor de goede zorgen.

Daarnaast hebben mijn twee andere promoteren, Egbert te Velde en Bart Fauser, een groot aandeel gehad. Egbert, door jou heb ik de fertiliteit leren kennen. Met je originele ideeën en je gedrevenheid voor onderzoek aan natuurlijke fertiliteit ben je een blijvende inspiratie voor mij. Bart, jij bent een stimulans voor me om door te gaan in dit vak. Jouw creativiteit, daadkracht en voortvarendheid zijn fenomenaal en zie ik als een groot voorbeeld voor mij. Ik dank jullie beiden voor jullie inzet en geduld.

De co-auteurs van al dan niet doorgezette artikelen dank ik voor hun constructieve commentaar. Met name Luc Smits, via wie ik heb kennisgemaakt met de database uit Quebec en Peter Pearson voor de goede discussies. Henri Leridon, I thank you for your participation in the paper that led to *Chapter 3*, and for your valuable comments on some of my ideas. I hope we will have the opportunity to work closer together in the future.

De meeste van de vrouwen die in het Canadese databestand voorkomen zijn inmiddels overleden. De grote aantallen kinderen die zij hebben gekregen maakten deze populatie geschikt voor het onderzoek in Deel I van het proefschrift. Zonder het te weten en zich schikkend naar het keurslijf van de religieus-sociale omgeving waarin

zij leefden, zijn zij belangrijk geworden voor de wetenschap. De grootsheid van hun prestatie is niet in woorden te vatten. Je restreints mes mots de grâce –ne signifiant aucunement avec moins d'appréciation- aux employées de l'Institut Interuniversitair de Recherches sur les Populations (IREP) à Chicoutimi, Québec Canada pour avoir mis à la disposition des données. Spécialement je remercies le directeur Gérard Bouchard, 'Super' Mario, et Michèle Jomphe pour ses e-mails toujours serviables et affectueux (sur des femmes qui peuvent appartenir à trois generations en même temps: 'God this is hard to explain in English and moreover by email!').

Het tweede en tevens grootste deel van dit proefschrift zou er nooit zijn gekomen zonder twee mensen: Babek Imani en Annemarie Mulders. Babek, de unieke database die er nu ligt is zowel in opzet als in inhoud grotendeels jouw werk geweest. Ik dank je voor de prettige samenwerking en voor je vriendschap. Annemarie, jij bent uniek in je hartelijkheid en warmte, en daarnaast een enorme volhouder. Waardering, en dank voor het werk dat je voor mij hebt gedaan. Daarnaast wil ik zeker noemen, de andere leden van de Rotterdamse fertiliteitsgroep, met wie ik al zovele jaren samen optrek en van wie ik steeds weer nieuwe dingen leer. Met name Joop Laven en Nick Macklon, meer recent Evert van Santbrink, de 'follicle girls en boys': Femke, Mark, Nicole, Christien, en eerder Diederick, Elena en Esther van het lab, ik maak met jullie allemaal steeds genoeglijke en tegelijkertijd leerzame momenten mee. En wie ik zeker ook wil noemen zijn Marita op het secretariaat en Lizka die zoveel weet te coördineren. Dank dat ik met jullie mag werken, en voor de hartelijkheid waarmee dat altijd gaat.

Claudine, ik bewonder je vasthoudendheid en je instelling om overal het positieve van in te zien. Het leek vanzelfsprekend dat ik eerder zou promoveren dan jij, hoewel ik daar zelf lang aan heb getwijfeld. Nu is het dan toch zover. In mijn eindsprint heb ik soms te weinig tijd en aandacht gehad voor jouw voortgang, terwijl dat juist zo hard nodig is in de taaie materie waar jij mee worstelt, mijn excuses daarvoor. Ik hoop het tijdens jouw eindsprint te kunnen goedmaken; gaan we er voor?

De vele samenwerkingen op fertiliteitsgebied nu en in het verleden, die mij stuk voor stuk hebben verrijkt, en waarvan de weerslag ongetwijfeld ook in dit proefschrift terug te vinden is. Ik noem allereerst Jan-Willem en Pieternel. Wat een fijne mensen zijn jullie allebei, en wat een prestatie dat het OFO-project gaat lukken: mooie promoties gaan dat worden. Ik betrek hierin meteen ook de rest van de OFO-project-groep, Fulco van der Veen, Peter Hompes en Patrick Bossuyt, maar bovenal Ben-Willem Mol: over drijvende kracht gesproken... BW, ik hoop dat we elkaar nog vaak zullen vinden in het fertiliteitsveld. Jij bent iemand om wie je moeilijk heen kunt ('niet te passeren, die man...'). Niet minder belangrijk voor mij is het 'minimal stimulation' project. Esther, met name door jouw vele talenten en je grote inzet is het een heel goed lopende studie geworden. Ook Cora en Suzanne hebben daar zeker hun aandeel in. Ik ben ontzettend benieuwd hoe dit verder zal lopen. Didi ('voordat we beginnen met de vergadering: René, hoe gaat het met je boekje?'), Bea met wie ik inmiddels stad en land heb afgereisd, en Rianne, Clazien en Leona: ook het project 'indicaties voor IVF' gaat iets heel groots worden. Ik prijs me gelukkig met jullie allen te mogen werken.

De collega's van MGZ. Allereerst wil ik Ewout Steyerberg noemen. Ewout, wij werken al zo lang samen, tegenwoordig minder dan voorheen als de siamese tweeling waar we soms voor werden aangezien. Ik heb ongelooflijk veel van je geleerd, en veel daarvan is in dit proefschrift terechtgekomen, dank, dank, dank.

Van je kamergenoten moet je het hebben. Pieta, met wie ik wel bijna getrouwd lijk, zo lang samen en zo vertrouwd: ik kon altijd m'n verhaal even kwijt, al ging het dan ook nergens over (whuu!) en Cecile: je hebt me in korte tijd meer gegeven dan je zelf voor mogelijk had gehouden. Het was niet minder gezellig met de oud-kamergenoten. Judith (lief en leed delen...), daar voor Margriet, en wie ik nooit zal vergeten: Veerle. Allen bedankt voor de nooit aflatende interesse en de gezelligheid en wat al niet.

De ondersteunende diensten van het instituut. Jullie zijn onmisbaar, maar worden niet altijd op waarde geschat. Met wie ik te maken heb gehad en nog steeds heb: Ton, Kees en Peter-altijd-bereid–een-brandje-te-blussen, bedankt. Op het secretariaat met name de secretaresses van Dik: Mirela, en daar voor Edith ('say no more..', zie hieronder) en weer daar voor Tineke. En daar ergens tussendoor en af-en-toe of juist heel vaak: Sonja. Mona voor de literatuur en voor nog zoveel meer, lieve Mona...bedankt! Tenslotte: Barry, Jolande en Selma. Het was en is voor jullie niet makkelijk met mijn lappendeken aan subsidies ('welk project is dit nou? Iets met fertility..'). Jo, ik ben zo blij dat ook voor jou de zon weer schijnt.

Mijn paranimfen Jan en Edith (zie hierboven). Jan, het voelt weldadig te weten dat mijn beste vriend straks naast me zal staan. Toch het samen fietsen maar weer oppakken, lijkt me? En Edith, ik leerde je kennen in de tijd van het gouden trio Mona, Sonja, Edith. Onvergetelijk. Ik ben blij dat we zo close zijn gebleven, ook na die tijd.

De vrienden uit Delft waren en zijn belangrijk: Ellen, Henk, Jaap, Wim, Arjo, Nico en Nicole. Bedankt voor de steun, de belangstelling en voor het gevoel een tweede familie te hebben. En dat jullie me serieus zijn blijven nemen, ook toen ik tussen 'die medici' ging werken.

Tot slot enkele woorden aan de mensen die mij het meest na staan. Mijn ouders, zus en broer. Het leek altijd zo ongrijpbaar wat ik aan het doen was. Nu zien jullie het resultaat. Papa en mama, ik ben nog een heel eind gekomen, wat nooit zou zijn gebeurd zonder jullie steun.

Laatst en liefst, Litanja. Door een toeval kwam jij in mijn leven, hoewel ik niet uitsluit dat ook hier een sturende kracht heeft meegespeeld. Wij hebben het zo goed samen, laten we samen verder gaan...

Curriculum vitae

René Eijkemans was born on May 14, 1963 in St Michiels Gestel. In 1981, he obtained the diploma Gymnasium beta at Gymnasium Bernrode in Heeswijk-Dinther. He then moved to Delft to start the study Mathematics at the Delft University of Technology. In 1991 he graduated in Mathematics, within the field of Differential Equations, after having completed his MSc study on 'the supersonic conical flow around a pyramid'.

That same year he joined the Center for Clinical Decision Sciences (head: Prof.dr. Dik Habbema), which at present is part of the Department of Public Health, ErasmusMC Rotterdam. There he started his career as a researcher with two clinical cost-effectiveness projects: 'Oncological follow-up of patients operated for colorectal cancer' and 'The value of ancillary diagnostics in dementia'. These projects were followed by other cost-effectiveness studies, on 'The evaluation of staging procedures of malignant lymphoma' and 'An individualized treatment protocol for chronic Hepatitis C'. At that time he became involved in fertility studies, first together with Prof.dr. Egbert te Velde from Utrecht, and later on also with Prof.dr. Bart Fauser from the fertility unit in Rotterdam. Since 1998 he has been working as a statistical consultant for the Rotterdam group, and has been involved in a number of projects on fertility. Currently he is participating in 3 ongoing projects funded by the ZON-MW special program for cost-effectiveness in fertility medicine.