
G. DILLARD
OMNE VIVUM EX OVO
E. Boispeau

Human fecundity under natural conditions and during in vitro fertilization

Human fecundity
under natural conditions
and during in vitro fertilization

Een wetenschappelijke proeve
op het gebied van de Medische Wetenschappen

Proefschrift ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op donderdag 6 februari 1997,
des namiddags om 1.30 uur precies

door

Anthonia Maria Stolwijk
geboren op 13 juni 1965 te Gouda

Promotores: Prof. dr. ir. G.A. Zielhuis
Prof. dr. A.L.M. Verbeek
Prof. dr. F.J.M. Gabreëls

Co-promotor: Dr. C.J.C.M. Hamilton, Bosch Medicentrum, Den Bosch

Manuscriptcommissie: Prof. dr. R.C.A. Sengers
Prof. dr. J.M.W.M. Merkus
Prof. dr. ir. J.D.F. Habbema, Erasmus Universiteit Rotterdam

Several studies presented in this thesis were financially supported by: 'Maria Roepaan', Centre for the Mentically Handicapped, Ottersum, the 'Van Walree fonds' of the Royal Dutch Academy of Sciences (KNAW), and 'The Netherlands Organization for Scientific Research' (NWO), The Netherlands.

Publication of this thesis was financially supported by: Organon Nederland bv.

ISBN 90 373 0356 0

Cover: G. Dillard naar de sculptuur "Omne vivum ex ovo" van E. Boisseau
(G.-J. Witkowski, La Génération Humaine. Paris: Georges Steinheil, 1886).

Ter herinnering aan mam

Contents

1. Introduction

Ways and means of regulating human fertility have expanded enormously during the second half of the 20th century. Contraceptives have been developed to prevent pregnancy (the most impressive being oral contraceptives), while assisted reproductive techniques have become available to improve the probability of pregnancy for subfertile couples. The ability to regulate fecundity and changes in society in the past few decennia have led to an increase in the age at which a woman gives birth to her first child. In the Netherlands, the average age in 1950 was 26.6 years; this decreased to 24.8 years in 1970 and was followed by an increase in 1990 to 27.6 years.[1,2] This trend can also be seen in other western societies.[3,4] A higher proportion of couples will seek professional help, because more problems with fecundity appear with increasing age[4] and the age span for reproduction is smaller if a woman decides to have her first baby at an older age. A Danish study performed in 1989 revealed that about 12% of women aged 25-44 years had sought medical advice at some time because of fertility problems. More than half of these women conceived spontaneously afterwards.[3] A Dutch study estimated that 10.4% of women will seek specialist medical help for infertility at some time during reproductive life (at age of 15-44 years) and that requests for advice from general practitioners will be 1.8 times higher.[5] A study performed in 1988 in the United States revealed that per year, about 2.3% of women of reproductive age received medical advice or treatment for infertility.[6] The use of medical assistance depends on factors such as the prevalence of fertility problems, the availability of infertility treatment for a population at a specific time, the cost and financial resources. Despite the advances in infertility treatments, three to four per cent of all women remain involuntarily childless.[7,8]

Greater knowledge about human fecundity is required to assist rational decision-making on prevention strategies and treatment modalities for infertility. The aim of this thesis is to increase the level of insight into the variation in human fecundity under natural conditions and during treatment with in vitro fertilization. Fecundity is defined as the ability to conceive within one menstrual cycle when pregnancy is desired and no contraceptive methods are used. Fecundability is the probability of conceiving during one such menstrual cycle.

Natural fecundity; The menstrual cycle

The central theme of the first part of this thesis is the fecundity of couples under natural conditions. We focus on women. Normal fecundability is estimated to be between 20% and 35% per cycle in women of 20-30 years of age.[1] An overview of the regulation of the menstrual cycle, from the initiation of follicular growth until ovulation, is given below.[9]

Primordial follicles are formed in the female fetus at a gestational age of 16-20 weeks. Each follicle contains an oocyte arrested in the diplotene stage of the first meiotic prophase. Follicular growth is a continuous process, occurring at all ages until the number of primordial follicles is exhausted. Follicular growth is initiated independently of hormonal influence. The number of developing follicles is proportional to the number of primordial follicles left.

During the menstrual cycle, a rise in follicle stimulating hormone (FSH) stimulates a group of follicles to further growth. FSH induces granulosa cells to aromatize androgen to oestrogen. The presence of FSH and oestrogen increase the FSH receptor content of the follicle. The dominant follicle is selected during days 5-7 after the first day of the menstrual cycle, when peripheral levels of oestradiol begin to rise and, because of negative feedback to the pituitary, peripheral FSH decreases. The number of FSH receptors on the granulosa cells and the vascularization of the theca cells of the dominant follicle are larger than those of the other follicles. Increased vascularization promotes the delivery of gonadotrophins to the dominant follicle. With declining FSH levels, the midfollicular rise in oestradiol was thought to give positive feedback on luteinizing hormone (LH) secretion.[9] However, a recent study has suggested that the increase in LH secretion is the consequence of a decline in the inhibitory effect of oestradiol when oestradiol levels decrease.[10] LH stimulates thecal cells to produce androgens which are converted into oestrogens through FSH-induced aromatization in the granulosa cells. LH initiates luteinization and progesterone production in the granulosa cells. The rise in progesterone decreases the inhibitory effect of oestrogen.[10] A series of enzymatic processes starts which eventually leads to follicle rupture. Ovulation occurs 34-36 hours after the onset of the LH surge. LH concentration must be maintained at a certain threshold level for 14-27 hours in order to achieve full oocyte maturation. The LH surge initiates the resumption of meiosis in the oocyte. In the normal cycle, the time from the LH midcycle surge to menses is consistently 14 days, although a luteal phase of between 11-17 days can also be considered normal. Wide variability in cycle length between women is due to the varying number of days required for follicular growth and maturation in the follicular phase.

Subfecundity under natural conditions

In the developed countries, menarche occurs at around the age of 13 years[11-15] and menopause at around the age of 51 years.[16-21] Menopause occurs because the supply of primordial follicles is depleted.[22] Follicular growth is then accelerated and fewer follicles per cycle grow.[22,23] The menstrual interval shows great variation and higher median values for seven years after menarche and for eight years before menopause.[24] In women with regular menstrual cycles who are younger than 20 years, or 40 years and older, decreased ovarian function might be expected because of diminished progesterone levels.[25] The proportion of anovulatory cycles decreases from 60% in women aged 12-14 years, to 5% in women aged 26-30 years and increases after the age of 40. The proportion of cycles with a shortened luteal phase shows the same trend.[26] It is more accurate to study changes in ovulatory function in relation to gynaecological age (i.e. years after menarche) than calender age. During the first 5 years after menarche, a high incidence of anovulatory cycles was observed.[27,28] The proportion of anovulatory cycles increased after a gynaecological age of about 35 years,[27] which is in agreement with the high incidence of anovulatory cycles in perimenopausal women who experienced a sudden break in menstrual cyclicity after many years of regular cycles.[29]

Besides the phases directly after menarche and before menopause, there can be other situations during reproductive life in which a woman is in a suboptimal state for reproduction. Such phases may occur just after stopping the use of oral contraceptives, just after pregnancy, during and after lactation, in relation with an endocrine disease, in the case of extreme under- or overweight and with weight changes.

After discontinuing the use of oral contraceptives, there may be an unintended effect in the form of prolonged interference with hormone production by the hypothalamus-pituitary-ovarian axis during the first few menstrual cycles. Reported consequences are anovulatory cycles[30,31] and a prolonged time to pregnancy.[32-34] The delay in conception was longer in the women who had used oral contraceptive pills that contained a higher dose of oestrogen.[34]

In the women who did not breast-feed, ovulation returned 40-50 days postpartum, while in the women who did breast-feed, the mean time to the first ovulation was 30-40 weeks.[35] Reproductive problems are also common in situations involving endocrine diseases such as diabetes mellitus, thyroid dysfunction and chronically increased or decreased gonadotrophin levels.[9] Generally, no ovulation occurs in these situations. In most of these disorders, reproductive capacity can largely be regained after adequate treatment.[9]

Ovarian function is also impaired in the case of extreme under- or overweight. In

obese women, the androgen:oestrogen ratio alters, which affects the feedback system, increases LH release and results in disrupted follicular maturation and anovulatory cycles.[36] The amount of visceral fat seems to play an important role in this hyperandrogenic situation.[37] The risk of ovulatory infertility was found to be the highest in obese women [Quetelet Index (QI) ≥27 kg/m^2]; it was also slightly increased in moderately overweight women (QI 25-26.9 kg/m^2) and in underweight women (QI <17 kg/m^2).[38] Lower luteal progesterone levels were even found in women of normal weight who had lost some weight by diet changes (on average 1.9 kg/month). According to the authors, this reflected a high proportion of anovulatory cycles.[39] In another study, normal-weight women had lost an average of 2.1 kg after a 3-day fast during the midfollicular phase. However, no impaired follicle growth or anovulation were observed by ultrasound, although LH pulses and LH production were diminished.[40] In normal-weight women who had lost an average of 4.9 kg during one menstrual cycle by following a vegetarian diet, ultrasound showed an increase in anovulatory cycles. In the diet cycle, LH pulses and concentrations were decreased during the follicular phase, while lower levels of oestradiol and progesterone were found during the luteal phase.[41]

Poor reproductive outcome under natural conditions of subfecundity

It has been hypothesized that women with diminished fecundity as in conditions mentioned above, have an increased risk of reproductive failures, for example spontaneous abortion, stillbirth and congenital malformations.[42] Several studies support this hypothesis. For instance, more early pregnancy losses,[43] spontaneous abortions[44-46] and premature infants[45] were found in subfecund women than in fecund women. An increased prevalence of infants with a very low birth weight was observed in women of 18 years or younger.[47] Even in a 'low risk group' of women (white, married, with age-appropriate education and receiving adequate prenatal care), young maternal age was related to an increase in low birth weight, prematurity and small for gestational age infants. The highest risk was found in women aged 13-15 years, while even in women aged 18-19 years, the risk was elevated in comparison with women aged 20-24 years.[48] However, most studies that addressed young maternal age and poor reproductive outcome did not show a relationship after adjustment for the confounding effects of factors such as socio-economic status and smoking.[49] In one study, there was no relationship with young reproductive age either after adjusting for confounding effects, but it was shown that there might be an increased risk of premature delivery in women who conceive within 2 years after menarche.[50] In women who were inseminated with donor semen, there was not only a decreased probability of conception after the age of 31, but also an increased risk of spontaneous abortion.[51] A

higher risk of spontaneous abortion after achieving a clinical pregnancy was observed with increasing age in subfertile women treated with ovulation induction and if necessary, inseminated with donor semen.[52] The incidence of chromosomal aberrations and genetic mutations in oocytes or spermatozoa increased with the age of the woman[53] or man.[54] Both at a young (<20 years) and old (≥40 years) maternal age, relatively high prevalences of congenital malformations and non-chromosomal congenital malformations in particular, were observed in liveborn children.[55] Women who were using oral contraceptives during pregnancy had an increased risk of chromosomal abnormalities in spontaneously aborted fetuses.[56,57] However, a meta-analysis did not show an increase in births with congenital malformations in women who had used oral contraceptives after their last menstrual period.[58]

Season as a natural cause of subfecundity

Another factor that may influence human fecundity is the season.[42] Most mammals show obvious seasonal variation in reproduction. The mechanism is apparently triggered by photoperiodicity. Changes in the photoperiod are registered by the retina and pulses are transmitted via the suprachiasmatic nucleus to the pineal gland.[59] The pineal gland produces melatonin which inhibits or stimulates the hypothalamus to activate ovulation, depending on the species. It is unclear whether the pineal gland and melatonin play a role in gonadal function in humans.[9] Hypothetically, remnants of a seasonal influence can be expected to exist in humans.

Several studies detected seasonal variation in human fecundity by means of measuring levels of hormones produced by pituitary and ovary,[60-62] menstrual interval length,[63] ovulation,[64,65] sperm characteristics,[66,67] the probability of becoming pregnant,[68,69] the time to pregnancy[70] and birth rates.[71-73] Seasonal patterns in reproductive failure were also detected in the form of early pregnancy loss,[74] spontaneous abortions[75-78] and congenital malformations.[79-81]

Requirements for studies on natural conditions of subfecundity

Some of the factors mentioned above which influence natural human fecundity have been studied extensively, for instance high maternal age, while others have not. New studies should focus on the latter factors. Such studies should follow a large population of women prospectively starting from the time they start trying to conceive. As the incidence of poor reproductive outcome is low, associations can only be discovered if data are obtained from a large number of women. If women are enroled in a study after they have started trying to conceive, subfecund women will be overrepresented and any early losses may go

undetected. A more realistic study design is a case-control study. However, such a study will be seriously biased due to a lack of valid data on the determinants, especially if weight change is the determinant under study. The effects of a preceding pregnancy and lactation on poor reproductive outcome are difficult to study, because most women do not want to become pregnant immediately following a pregnancy. The residual effects of oral contraceptives have become less relevant, because the dose of hormones has diminished over the years. In contrast to the studies on the above-mentioned factors, studies on the impact of the season on fecundity might give additional information about the variation in human fecundity. The first part of this thesis presents several studies on seasonality in human fecundity and reproductive failure.

In vitro fertilization: The technique

For a better understanding of human fecundity and to assist decisions on infertility treatment, is it worthwhile to investigate the variation in fecundity in situations in which explicit attempts are made to increase fecundity, such as during treatment with assisted reproductive techniques. In vitro fertilization with embryo transfer (IVF) is one of the most widely applied techniques. Data from IVF treatments might help to identify factors that have predictive value for fecundity.

In 1978, the first child was born after IVF.[82] Initially, IVF was used in women with tubal disease. Later the treatment was also offered to couples with a male infertility factor, unexplained infertility, endometriosis and an immunological cause for infertility.[83] However, it has not yet been demonstrated whether IVF is effective in these cases.[84] In an attempt to improve fertilization, various techniques were developed, for example subzonal insertion of sperm (SUZI).[85] In 1992 a new technique, intracytoplasmatic sperm injection (ICSI) was introduced[86] with promising results. If necessary, donated spermatozoa and oocytes can be used. Other assisted reproductive techniques such as gamete intrafallopian transfer (GIFT)[87] and zygote intrafallopian transfer (ZIFT)[88] have also been developed, but are not widely used any more. The choice of technique depends mainly on the infertility factors of the couple, the experience of the medical centre and the financial cost.

The IVF treatment method differs between clinics. In general,[9,83] IVF treatment starts with ovarian stimulation using hormones, to achieve the simultaneous growth of many follicles. The natural hormonal production is often suppressed by a gonadotrophin releasing hormone agonist (GnRH-a). The dosage and timing of administration vary.

Human menopausal gonadotropin (hMG) and/or FSH are often used to stimulate follicular growth. When several large follicles are present, human chorionic gonadotropin (hCG) is administered to mimic the LH surge. Oocyte retrieval is performed 34-38 hours afterwards, mostly transvaginally under ultrasound guidance. The oocytes are inseminated by a portion of the semen, usually several hours after follicle aspiration. Semen has to be prepared in some way before it can be used for insemination. In the case of severe male subfertility, ICSI can be used as a fertilization technique. Two or three days after insemination, embryos are selected for transfer. The luteal phase can be supplemented with progesterone or hCG. A pregnancy test (hCG test) can be performed after 11 days following oocyte retrieval. An ongoing pregnancy has been achieved if the pregnancy continues to exist for longer than 12 weeks after oocyte retrieval and fetal heart activity can be detected by ultrasound.

Fecundity during treatment with in vitro fertilization

It is important to be able to predict the success of treatment with in vitro fertilization, because the technique is expensive and the emotional and physical demands are high. The cost per IVF treatment cycle varies worldwide. In the Netherlands, one IVF treatment costs about 3,000 to 3,500 Dutch gilders (i.e. about $ 1,800 to 2,000). The average price per IVF cycle in the United States of America is about $7,000 to $11,000.[89] Additionally, the cost of prenatal care, neonatal care and method of delivery in IVF pregnancies are higher than in natural pregnancies, especially because of the higher probability of multiple gestations.[89-91] The goal of patients who undergo IVF is to achieve a liveborn child. The likelihood of achieving this goal should be predicted at the start of treatment and at every successive stage during the treatment, e.g. after an unsuccessful cycle. Most studies on the success of IVF treatment considered only a few factors at the same time. For instance, poor IVF results have been found among older women,[92-94] in couples with a male infertility factor[93 95] and in women with a high basal FSH level.[96,97] However, various factors may have a simultaneous effect on the IVF result. For the clinician who is counselling a patient, the univariate relations will not always contribute to a clear prognosis. Therefore, to make a reliable prognosis, the predictive value of all factors should be considered in one prognostic model.

In the second part of this thesis, data of IVF treatments since 1989 are used to detect factors that have predictive value regarding the success of fertility treatment. Data were obtained from three clinics in the Netherlands and one in the United States of America. In particular, the effect of the woman's age on IVF outcome is studied, a topic closely related to the study of variation in fecundity under natural conditions.

Contents of the thesis

The central goal of this thesis is to study human fecundity. Two different sets of circumstances were evaluated: natural conditions and those during medical treatment for subfertility. Within this context, specific choices were made. Above, arguments are presented for addressing two core questions: 1) Is there seasonal variation in human fecundity? and 2) Which combinations of factors are of prognostic significance for the results of IVF treatment?

The first question is addressed in Chapter 2. Several methodological aspects of the study on seasonality are described: the choice of a parameter to describe seasonal patterns in reproductive outcome, a method to test seasonality and the impact of bias caused by seasonality in pregnancy planning on poor reproductive outcome (Chapter 2.1). A study on whether there is seasonal variation in the occurrence of ovulatory cycles is described in Chapter 2.2. A side step is taken to evaluate seasonal fecundity in the case of IVF treatment (Chapter 2.3). Two studies are performed to study seasonal patterns in the time to pregnancy (Chapter 2.4). As seasonal influence on hormone production by the hypothalamus-pituitary-ovarian axis may lead to problems in the completion of the first meiotic division, which occurs just before ovulation, the literature on seasonality in the prevalence of Down syndrome at birth is reviewed (Chapter 2.5).

The studies described in Chapter 3 focus on the second question regarding prognostic factors for the results of IVF treatment. In paragraph 3.1, the cumulative pregnancy rate after successive treatments with IVF is computed in several ways; this shows the importance of the underlying assumptions. Prognostic models for the probability of achieving an ongoing pregnancy after IVF treatment are developed and tested (Chapters 3.2 and 3.3). Ongoing pregnancy is used as an indicator for the occurrence of at least one live birth. There were two reasons for this: the better availability of information about ongoing pregnancy in comparison with information at birth and the close relationship between the two results. Another study is performed to gain more insight into the importance of the woman's age on the success of IVF by using information about IVF cycles with donated oocytes (Chapter 3.4).

In the final chapter (Chapter 4), the information obtained in all the studies described in Chapters 2 and 3 is discussed in the light of the core element in this thesis, i.e. human fecundity. In addition, potential implications for clinical practice are discussed and recommendations for further research are given. The thesis ends with a summary.

References

1. Bonsel GJ, Van der Maas PJ. Aan de wieg van de toekomst. Scenario's voor de zorg rond de menselijke voortplanting 1995-2000. Houten/Diemen: Bohn Stafleu Van Loghum; 1994.
2. De Graaf A. In Nederland worden vrouwen laat moeder. Mndstat bevolk (CBS) 1992;40:16-18.
3. Schmidt L, Münster K, Helm P. Infertility and the seeking of infertility treatment in a representative population. Br J Obstet Gynaecol 1995;102:978-984.
4. Lansac J. Delaying parenting. Is delayed childbearing a good thing? Hum Reprod 1995;10:1033-1035.
5. Beurskens MPJC, Maas JWM, Evers JLH. Subfertiliteit in Zuid-Limburg: berekening van incidentie en van beroep op specialistische zorg. Ned Tijdschr Geneeskd 1995;139:235-238.
6. Mosher WD, Pratt WF. Fecundity and infertility in the United States: incidence and trends. Fertil Steril 1991;56:192-193.
7. Gunnell DJ, Ewings P. Infertility prevalence, needs assessment and purchasing. J Public Health Med 1994;16:29-35.
8. Schmidt L, Münster K. Infertility, involuntary infecundity, and the seeking of medical advice in industrialized countries 1970-1992: a review of concepts, measurements and results. Hum Reprod 1995;10:1407-1418.
9. Speroff L, Glass RH, Kase NG. Clinical gynecologic endocrinology and infertility. 5th ed. Baltimore, MD: Williams & Wilkins; 1994.
10. Levran D, Ben-Shlomo I, Luski A, et al. A reappraisal of the feedback effects of oestradiol upon luteinizing hormone surge. Hum Reprod 1995;10:3117-3120.
11. Roede MJ, Van Wieringen JC. Growth diagrams 1980. Netherlands third nation-wide survey. T Soc Gezondheidsz 1985;63 (Suppl.):1-34.
12. Rimpelä AH, Rimpelä MK. Towards an equal distribution of health? Socioeconomic and regional differences of the secular trend of the age of menarche in Finland from 1979 to 1989. Acta Paediatr 1993;82:87-90.
13. Dóber I, Királyfalvi L. Pubertal development in south-Hungarian boys and girls. Ann Hum Biol 1993;20:71-74.
14. Tryggvadóttir L, Tulinius H, Lárusdóttir M. A decline and a halt in mean age at menarche in Iceland. Ann Hum Biol 1994;21:179-186.
15. Veronesi FM, Gueresi P. Trend in menarcheal age and socioeconomic influence in Bologna (Northern Italy). Ann Hum Biol 1994;21:187-196.
16. Jaszmann L, Van Lith ND, Zaat JCA. The age at menopause in the Netherlands. The statistical analysis of a survey. Int J Fertil 1969;14:106-117.
17. Brand PC, Lehert PH. A new way of looking at environmental variables that may affect the age at menopause. Maturitas 1978;1:121-132.
18. Willett W, Stampfer MJ, Bain C, *et al.* Cigarette smoking, relative weight, and menopause. Am J Epidemiol 1983;117:651-658.
19. Stanford JL, Hartge G, Brinton LA, Hoover RN, Brookmeyer R. Factors influencing the age at natural menopause. J Chronic Dis 1987;40:995-1002.
20. McKinlay SM, Brambilla DJ, Posner JG. The normal menopause transition. Maturitas 1992;14:103-115.
21. Luoto R, Kaprio J, Uutela A. Age and natural menopause and sociodemographic status in Finland. Am J Epidemiol 1994;139:64-76.
22. Richardson SJ, Senikas V, Nelson JF. Follicular depletion during the menopausal transition: evidence for accelerated loss and ultimate exhaustion. J Clin Endocrinol Metab 1987;65:1231-1237.

23. Gosden RG. Follicular status at the menopause. Hum Reprod 1987;2:617-621.
24. Treloar AE, Boynton RE, Behn BG, Brown BW. Variation of the human menstrual cycle through reproductive life. Int J Fertil 1967;12:77-126.
25. Lipson SF, Ellison PT. Normative study of age variation in salivary progesterone profiles. J Biosoc Sci 1992;24:233-244.
26. Döring GK. The incidence of anovulatory cycles in women. J Reprod Fert 1969;Suppl. 6:77-81.
27. Vollman RF. The menstrual cycle. Major problems in obstetrics and gynecology, volume 7. Philadelphia: WB Saunders Company; 1977.
28. Metcalf MG, Skidmore DS, Lowry GF, Mackenzie JA. Incidence of ovulation in the years after the menarche. J Endocrinol 1983;97:213-219.
29. Metcalf MG, Donald RA, Livesey JH. Classification of menstrual cycles in pre- and perimenopausal women. J Endocrinol 1981;91:1-10.
30. Rice-Wray E, Correu S, Gorodovsky J, Esquivel J, Goldzieher JW. Return of ovulation after discontinuance of oral contraceptives. Fertil Steril 1967;18:212-218.
31. Pinkerton GD, Carey HM. Post-pill anovulation. Med J Aust 1976;1:220-222.
32. Linn S, Schoenbaum SC, Monson RR, Rosner B, Ryan KJ. Delay in conception for former 'pill' users. JAMA 1982;247:629-632.
33. Harlap S, Baras M. Conception-waits in fertile women after stopping oral contraceptives. Int J Fertil 1984;29:73-80.
34. Bracken MB, Hellenbrand KG, Holford TR. Conception delay after oral contraceptive use: The effect of estrogen dose. Fertil Steril 1990;53:21-27.
35. Wang IY, Fraser IS. Reproductive function and contraception in the postpartum period. Obstet Gynecol Surv 1994;49:56-63.
36. Sacks PC. The menstrual cycle. In: Scialli AR, Zinaman MJ, (Eds.). Reproductive toxicology and infertility. New York: McGraw-Hill, Inc. 6, The menstrual cycle. 1993;133-185.
37. Leenen R, Van der Kooy K, Seidell JC, Deurenberg P, Koppeschaar HP. Visceral fat accumulation in relation to sex hormones in obese men and women undergoing weight loss therapy. J Clin Endocrinol Metab 1994;78:1515-1520.
38. Grodstein F, Goldman MB, Cramer DW. Body mass index and ovulatory infertility. Epidemiology 1994;5:247-250.
39. Lager C, Ellison PT. Effect of moderate weight loss on ovarian function assessed by salivary progesterone measurements. American Journal of Human Biology 1990;2:303-312.
40. Olson BR, Cartledge T, Sebring N, Defensor R, Nieman L. Short-term fasting affects luteinizing hormone secretory dynamics but not reproductive function in normal-weight sedentary women. J Clin Endocrinol Metab 1995;80:1187-1193.
41. Pirke KM, Schweiger U, Strowitzki T, *et al.* Dieting causes menstrual irregularities in normal weight young women through impairment of episodic luteinizing hormone secretion. Fertil Steril 1989;51:263-268.
42. Jongbloet PH. The effect of preovulatory overripeness of human eggs on development. In: Blandau RJ, (Ed.). Aging Gametes. Basel: S. Karger AG, The effect of preovulatory overripeness of human eggs on development. 1975;300-329.
43. Hakim RD, Gray RH, Zacur H. Infertility and early pregnancy loss. Am J Obstet Gynecol 1995;172:1510-1517.
44. Rachootin P, Olsen J. Prevalence and socioeconomic correlates of subfecundity and spontaneous abortion in Denmark. Int J Epidemiol 1982;11:245-249.
45. Joffe M, Li Z. Association of time to pregnancy and the outcome of pregnancy. Fertil Steril

1994;62:71-75.
46. Basso O, Olsen J, Bisanti L, Juul S, Boldsen J, and the European Study Group on Infertility and Subfecundity. Are seasonal preferences in pregnancy planning a source of bias in studies of seasonal variation in reproductive outcomes? Epidemiology 1995;6:520-524.
47. Miller HS, Lesser KB, Reed KL. Adolescence and very low birth weight infants: a disproportionate association. Obstet Gynecol 1996;87:83-88.
48. Fraser AM, Brockert JE, Ward RH. Association of young maternal age with adverse reproductive outcomes. N Engl J Med 1995;332:1113-1117.
49. Goldenberg RL, Klerman LV. Adolescent pregnancy - another look. N Engl J Med 1995;332:1161-1162.
50. Mitchell LE, Bracken MB. Reproductive versus chronologic age as a predictor of low birth weight, preterm delivery and intrauterine growth retardation in primiparous women. Ann Hum Biol 1990;17:377-386.
51. Van Noord-Zaadstra BM, Looman CWN, Alsbach H, Habbema JDF, Te Velde ER, Karbaat J. Delaying childbearing: Effect of age on fecundity and outcome of pregnancy. Br Med J 1991;302:1361-1365.
52. Smith KE, Buyalos RP. The profound impact of patient age on pregnancy outcome after early detection of fetal cardiac activity. Fertil Steril 1996;65:35-40.
53. Plachot M, De Grouchy J, Junca A-M, Mandelbaum J, Salat-Baroux J, Cohen J. Chromosomal analysis of human oocytes and embryos in an in vitro fertilization program. Ann N Y Acad Sci 1988;541:384-397.
54. Bordson BL, Leonardo SV. The appropriate upper age limit for semen donors: a review of the genetic effects of paternal age. Fertil Steril 1991;56:397-401.
55. Croen LA, Shaw GM. Young maternal age and congenital malformations: a population-based study. Am J Publ Hlth 1995;85:710-713.
56. Carr DH. Chromosome studies in selected spontaneous abortions: 1. Conception after oral contraceptives. Can Med Ass J 1970;103:343-348.
57. Poland BJ, Ash KA. The influence of recent use of an oral contraceptive on early intrauterine development. Am J Obstet Gynecol 1973;116:1138-1142.
58. Bracken MB. Oral contraceptives and congenital malformations in offspring: a review and meta-analysis of the prospective studies. Obstet Gynecol 1990;76:552-557.
59. Hofman MA, Swaab DF. The human hypothalamus: comparative morphometry and photoperiodic influences. In: Swaab DF, Hofman MA, Mirmiran M, Ravid R, Van Leeuwen FW, (Eds.). Progress in brain research. Amsterdam: Elseviers Science Publishers B.V. 10, The human hypothalamus: comparative morphometry and photoperiodic influences. 1992;133-147.
60. Kauppila A, Kivelä A, Pakarinen A, Vakkuri O. Inverse seasonal relationship between melatonin and ovarian activity in humans in a region with a strong seasonal contrast in luminosity. J Clin Endocrinol Metab 1987;65:823-828.
61. Kivelä A, Kauppila A, Ylöstalo P, Vakkuri O, Lappäluoto J. Seasonal, menstrual and circadian secretions of melatonin, gonadotropins and prolactin in women. Acta Physiol Scand 1988;132:321-327.
62. Martikainen H, Ruokonen A, Tomás C, Kauppila A. Seasonal changes in pituitary function: amplification of midfollicular luteinizing hormone secretion during the dark season. Fertil Steril 1996;65:718-720.
63. Sundararaj N, Chem M, Gatewood L, Hickman L, McHugh R. Seasonal behavior of human menstrual cycles: a biometric investigation. Hum Biol 1978;50:15-31.
64. Timonen S, Franzas B, Wichmann K. Photosensibility of the human pituitary. Ann Chir Gynaec

Fenn 1964;53:165-172.
65. Rameshkumar K, Thomas JA, Mohammed A. Atmospheric temperature & anovulation in south Indian women with primary infertility. Indian J Med Res 1992;96:27-28.
66. Levine RJ. Seasonal variation in human semen quality. In: Zorgniotti AW, (Eds.). Temperature and environmental effects on the testis. New York: Plenum Press, Seasonal variation in human semen quality. 1991;89-96.
67. Levine RJ. Male factors contributing to the seasonality of human reproduction. Ann N Y Acad Sci 1994;709:29-45.
68. Paraskevaides EC, Pennington GW, Naik S. Seasonal distribution in conceptions achieved by artificial insemination by donor. Br Med J 1988;297:1309-1310.
69. Kallan JE, Udry JR. Demographic components of seasonality of pregnancy. J Biosoc Sci 1989;21:101-108.
70. Nonaka K, Desjardins B, Légaré J, Charbonneau H, Miura T. Effects of maternal birth season on birth seasonality in the Canadian population during the seventeenth and eighteenth centuries. Hum Biol 1990;62:701-717.
71. Lam DA, Miron JA. The seasonality of births in human populations. Ann Arbor, Michigan: Population Studies Center. University of Michigan; 1987.
72. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: I. Biology, sociology or both? J Biol Rhythm 1990;5:195-216.
73. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: II. Environmental correlations. J Biol Rhythm 1990;5:217-239.
74. Weinberg CR, Moledor E, Baird DD, Wilcox AJ. Is there a seasonal pattern in risk of early pregnancy loss? Epidemiology 1994;5:484-489.
75. Sandahl B. A study of seasonal and secular trends in incidence of stillbirths and spontaneous abortions in Sweden. Acta Obstet Gynecol Scand 1974;53:251-257.
76. Kallan JE, Enneking EA. Seasonal patterns of spontaneous abortion. J Biosoc Sci 1992;24:71-75.
77. Warren CW, Gwinn ML, Rubin GL. Seasonal variation in conception and various pregnancy outcomes. Soc Biol 1986;33:116-126.
78. McDonald AD. Seasonal distribution of abortions. Br J Prev Soc Med 1971;25:222-224.
79. Wehrung DA, Hay S. A study of seasonal incidence of congenital malformations in the United States. Br J Prev Soc Med 1970;24:24-32.
80. Bound JP, Harvey PW, Francis BJ. Seasonal prevalence of major congenital malformations in the Flyde of Lancashire 1957-1981. J Epidemiol Community Health 1989;43:330-342.
81. Liederman J, Flannery KA. Fall conception increases the risk of neurodevelopmental disorder in offspring. Journal of Clinical and Experimental Neuropsychology 1994;16:754-768.
82. Steptoe PC, Edwards RG. Birth after the reimplantation of a human embryo (letter). Lancet 1978;2:366.
83. Van den Eede B. Investigation and treatment of infertile couples: ESHRE guidelines for good clinical and laboratory practice. Hum Reprod 1995;10:1246-1271.
84. Buitendijk SE. Evidence-based in-vitro fertilisation (letter). Lancet 1995;346:901.
85. Ng S-C, Bongso A, Sathananthan H, Ratman SS. Micromanipulation: its relevance to human in vitro fertilization. Fertil Steril 1990;53:203-219.
86. Palermo G, Joris H, Devroey P, Van Steirteghem AC. Pregnancies after intracytoplasmatic injection of single spermatozoon into an oocyte. Lancet 1992;340:17-18.
87. Asch RH, Ellsworth LR, Balmaceda JP, Wong PC. Pregnancy after translaparoscopic gamete intrafallopian transfer (letter). Lancet 1984;ii:1034-1035.

88 Devroey P, Braeckmans P, Smitz J, *et al* Pregnancy after translaproscopic zygote intrafallopian transfer in a patient with sperm antibodies (letter) Lancet 1986,1 1329
89 Goldfarb JM, Austin C, Lisbona H, Peskin B, Clapp M Cost-effectiveness of in vitro fertilization Obstet Gynecol 1996,87 18-21
90 Callahan TL, Hall JE, Ettner SL, Christiansen CL, Greene MF, Crowley WF The economic impact of multiple-gestation pregnancies and the contribution of assisted-reproduction techniques to their incidence N Engl J Med 1994,331 244-249
91 Neumann PJ, Gharib SD, Weinstein MC The cost of a successful delivery with in vitro fertilization N Engl J Med 1994,331 239-243
92 Piette C De Mouzon J, Bachelot A, Spira A In-vitro fertilization influence of women's age on pregnancy rates Hum Reprod 1990,5 56-59
93 FIVNAT Evolutions des criteres pronostiques de fecondation in vitro selon le rang de la tentative Contracept Fertil Sex 1994,22 282-286
94 Padilla SL Prognostic indicators for in vitro fertilization success Maryland Medical Journal 1995,44 197-203
95 Haan G, Bernardus RE, Hollanders JMG, Leerentveld RA, Prak FM, Naaktgeboren N Results of IVF from a prospective multicentre study Hum Reprod 1991,6 805-810
96 Toner JP, Philput CB, Jones GS, Muasher SJ Basal follicle-stimulating hormone level is a better predictor of in vitro fertilization performance than age Fertil Steril 1991,55 784-791
97 Khalifa F, Toner JP, Muasher SJ, Acosta AA Significance of basal follicle-stimulating hormone levels in women with one ovary in a program of in vitro fertilization Fertil Steril 1992,57 835-839

2. Seasonal influence on fecundity

2.1. Methodology

2.1.1. Comparison of parameters to study seasonality in reproductive outcome

2.1.2. Studying seasonality by using a cosine function in regression analysis

2.1.3. Seasonality bias in poor reproductive outcome

2.1.1. Comparison of parameters to study seasonality in reproductive outcome

A.M. Stolwijk, G.A. Zielhuis, P.H. Jongbloet

This report compares validity, precision and convenience of three frequently used parameters used to describe seasonality in reproductive outcome: prevalence, index and ratio observed versus expected. Each has its advantages and disadvantages; the choice of the parameter depends on the aim of the study. It is advisable to report the number of cases and referents per month, because all three parameters can be derived from this information.

European Journal of Epidemiology 1995;11:355-358

Seasonality in reproductive outcome is a subject of epidemiological research because of suspected seasonally-bound etiological factors. The answer to the question of whether there is seasonal variation in the occurrence of a specific outcome, for instance births with congenital malformations, is only relevant if this seasonal pattern deviates from the seasonal pattern in the reference group, e.g. all births. Moreover, one might be interested in whether the observed pattern is comparable with the results of other studies.

No consensus exists regarding which parameter should be used to describe seasonality in reproductive outcome. Three frequently used parameters are: prevalence (for instance by Leck,[1] McDonald[2] and Stark & Mantel[3]), index (by Huntington[4] and Jongbloet & Vrieze[5]) and ratio observed versus expected (by Harlap,[6] Leisti *et al*,[7] Källén & Måsbäck[8]). This paper compares the performance of these parameters to describe seasonality in births with congenital malformations (CM).

The prevalence is expressed as the number of cases born with CM in month *m* relative to the number of births (B) in the same month, per 1,000 births. In order to conclude whether the prevalence in a specific month is low or high, this number should be compared to the mean CM prevalence (all cases with CM per 1,000 births).

$$\frac{CM_m}{B_m} \quad 1{,}000$$

The definition of index of CM is: the number of cases with CM born in each month, corrected for the number of days in that month (d_m), compared to the average number of births with CM a month. This has to be compared with the index of births in the same month.

$$\frac{\frac{CM_m}{d_m} \cdot \frac{365.25}{12}}{\frac{\sum_{m=1}^{12} CM_m}{12}}$$

A common method is to compare the observed number of births with CM in each month to the expected number of cases with CM born in the same month. The expected number of cases can be computed by multiplying the total number of births

with CM by the proportion of all births in the same month in relation to the number of births throughout the whole year.

$$\frac{CM_m}{\sum_{m=1}^{12} CM_m \cdot \frac{B_m}{\sum_{m=1}^{12} B_m}}$$

Three criteria are used to compare the three parameters: validity, precision and convenience.

Validity and precision of the three parameters were satisfactory because the monthly number of births with CM in comparison with the number of all births expressed as either a prevalence, or index, or ratio observed versus expected, could easily be transformed into each other:

from

$$\frac{CM_m \,/\, \sum_{m=1}^{12} CM_m}{B_m \,/\, \sum_{m=1}^{12} B_m}$$

to

$$\frac{Prevalence_m}{Mean\ prevalence} = \frac{(CM_m \,/\, B_m) \cdot 1{,}000}{(\sum_{m=1}^{12} CM_m \,/\, \sum_{m-1}^{12} B_m) \cdot 1{,}000} \qquad (1)$$

$$\frac{Index\ CM_m}{Index\ B_m} = \frac{(CM_m \,/\, d_m) \cdot (365.25/12) \,/\, (\sum_{m=1}^{12} CM_m \,/12)}{(B_m \,/\, d_m) \cdot (365.25/12) \,/\, (\sum_{m=1}^{12} B_m \,/12)} \qquad (2)$$

$$\frac{Observed_m}{Expected_m} = \frac{CM_m}{\sum_{m=1}^{12} CM_m \cdot (B_m \,/\, \sum_{m=1}^{12} B_m)} \qquad (3)$$

Thus, validity and precision are the same for all three parameters. However, comparison of the results of other studies which used different parameters can only be done if information is available on each component, i.e. the number of births with CM and the total number of births in each month.

A second comparison concerns the convenience of each parameter, i.e. the degree of insight that is provided into the results and the statistical properties of the parameter for studying seasonal differences.

Whether the parameters provide insight into the results can be illustrated using the data derived from a study on seasonality in Down syndrome (DS) among live births in Quebec from 1958 to 1967.[2] The number of DS and live births for each month are given in the Table 1, together with the computed prevalence, indexes, and ratio observed versus expected.

Table 1. Monthly number of DS and live births* and computed parameters

Month	Number of DS births	Number of live births	Prevalence	Index		Ratio observed versus expected
				DS births	Live births	
January	215	105,858	2.03	1.06	0.97	1.09
February	174	97,678	1.78	0.94	0.98	0.96
March	223	114,093	1.95	1.10	1.04	1.05
April	217	114,625	1.89	1.10	1.08	1.02
May	213	118,051	1.80	1.05	1.08	0.97
June	179	113,511	1.58	0.91	1.07	0.85
July	215	112,819	1.91	1.06	1.03	1.03
August	205	107,752	1.90	1.01	0.98	1.02
September	210	108,773	1.93	1.07	1.03	1.04
October	182	103,792	1.75	0.89	0.95	0.94
November	183	96,844	1.89	0.93	0.91	1.02
December	182	96,448	1.89	0.89	0.88	1.02
Total	2,398	1,290,244	1.86	1.00	1.00	1.00

* Data about monthly DS and live births from McDonald.[2]

For descriptive epidemiology, the monthly prevalence can be reported. For instance, 'In the month January, on average 2 DS cases were born among 1,000 live births in Quebec during the period 1958-1967'. The other two parameters do not have any meaning on their own and have to be reported together with the other monthly indexes and ratios.

For revealing the kind and the amount of deviation, the parameters can be criticized on the meaning of the figures and on the information after visualization. Monthly DS prevalences do not reflect deviations; for that purpose they have to be compared to the mean monthly DS prevalence. Monthly indexes can be interpreted as the percentage of deviation from the mean. For instance, the DS index in January was 1.06 and can be interpreted as a surplus of 6 per cent DS births in comparison with the average number of DS births a month. In the same month, the index of live births was 0.97, which means a deficit of 3 per cent in comparison with the average number of live births a month. Whereas these two indexes have to be compared before it can be concluded whether the number of DS births is higher than expected based on the number of live births in that month, the ratio observed versus expected reflects this deviation at once. In January, the ratio was 1.09 which means that more DS cases were born in January than were expected, based on the number of live births in that month.

The three parameters over a period of a year are plotted in Figures 1-3. The DS prevalence in Figure 1 and the ratio observed versus expected in Figure 3 clearly show the month in which DS births were higher or lower than average. In November-January, March-April and July-September the prevalences were higher than the mean prevalence and the ratios observed versus expected were higher than one. In contrast, for interpreting the indexes, the two indexes in each month must be compared. During the months mentioned above, the indexes of DS were higher than for live births. However, an advantage of the indexes (Figure 2) over the other two parameters is that seasonality in DS births can be examined together with seasonality in live births.

The statistical properties of the three parameters differ widely. An advantage of prevalence is that confidence limits can be computed for revealing the precision of the point estimate. Two monthly prevalences will be different if their confidence intervals do not overlap. Furthermore, prevalences per month can be compared between studies and pooled prevalences can be computed if several studies have the same design.[9] For indexes, no statistics are available for comparison within a study or between studies. The ratio observed versus expected per month cannot be compared directly between studies. However, a chi-square test can be performed to test for differences in frequency between months within a study.[10]

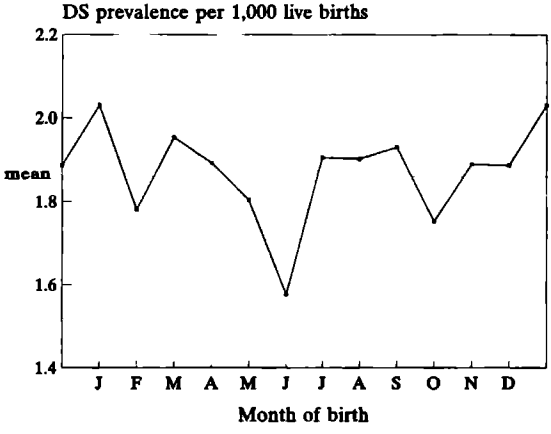


Figure 1. Prevalence

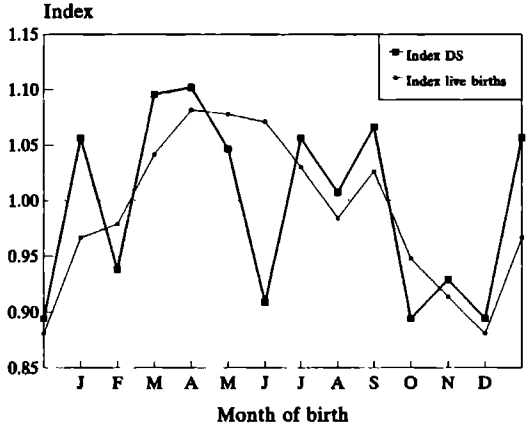


Figure 2. Index

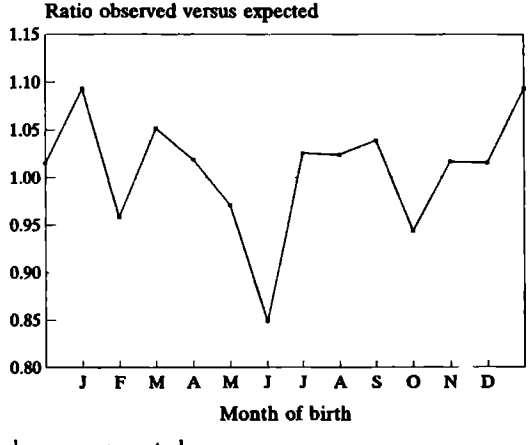


Figure 3. Ratio observed versus expected

All three parameters allow for adjustment for confounding by means of stratification. Moreover, if logistic regression analysis is used to adjust for confounding (for instance with the months included into the model as 11 dummy variables), the results could be expressed as adjusted prevalences. Such a method of multivariable analysis has advantages over stratification, in particular when dealing with small numbers per stratum and with many confounders. If some confounders are present in a study for seasonality in reproductive outcome, the presentation of stratified crude data is recommended, which is a requisite to allow for comparison between studies.

In this article only three parameters to describe seasonality were compared. The choice of a statistical test for studying seasonality is not be discussed here; for an overview see Reijneveld.[11] Moreover, other aspects in studies on seasonality in reproductive outcome may be important for the validity of the study. For instance, seasonality in a specific type of CM at birth may be observed only because of a relation of this type of CM with preterm birth. Also other factors may confound a seasonal pattern, for instance if pregnancy planning varies between age groups and social economic levels. In addition, a seasonal pattern of CM in births may differ from a seasonal pattern of CM when all conceptions are used as reference, because of a seasonal pattern in pregnancy loss. A study of the occurrence of CM in conceptions must be performed prospectively, testing all women who are trying to become pregnant each month as the effect of early pregnancy loss will be underestimated otherwise. Detecting CM in aborted fetuses is practically impossible. But even if the presence of CM in abortuses can not be determined, the description of seasonality in total pregnancy loss can facilitate the interpretation of the pattern of CM at birth.

In summary, the choice of a parameter to study seasonality in reproductive outcome depends on the aim of the study. Monthly prevalences give a direct reflection of reality and can be reported in isolation. Prevalences are suitable for comparing data within and between studies. The ratio observed versus expected can be used to draw conclusions within a study. An advantage of this ratio is that the figure shows the amount and kind of deviation in the number of cases in a month. The index is the best parameter only if the seasonality in cases as well as the seasonality in the reference group are of interest. As all three parameters can be derived from the monthly number of cases and referents, it is advisable to report this specific information, instead of the monthly parameter alone.

References

1. Leck I. Incidence and epidemicity of Down's syndrome. Lancet 1966;2:457-460.
2. McDonald AD. Yearly and seasonal incidence of mongolism in Quebec. Teratology 1972;6:1-3.
3. Stark CR, Mantel N. Lack of seasonal- or temporal-spatial clustering of Down's syndrome births in Michigan. Am J Epidemiol 1967;86:199-213.
4. Huntington E. Season of birth. Its relation to human abilities. New York: John Wiley & Sons, Inc. 1938.
5. Jongbloet PH, Vrieze OJ. Down syndrome: Increased frequency of maternal meiosis I nondisjunction during the transitional stages of the ovulatory seasons. Human Genet 1985;71:241-248.
6. Harlap S. A time-series analysis of the incidence of Down's syndrome in West Jerusalem. Am J Epidemiol 1974;99:210-217.
7. Leisti J, Vahtola L, Linna SL, Herva R, Koskela SL, Vitali M. The incidence of Down syndrome in northern Finland with special reference to maternal age. Clin Genet 1985;27:252-257.
8. Källén B, Måsbäck A. Down syndrome. Seasonality and parity effects. Hereditas 1988;109:21-27.
9. Rothman KJ. Modern epidemiology. Boston/Toronto: Little, Brown and Company; 1986.
10. Siegel S, Castellan NJ. Nonparametric statistics of the behavioral sciences. 2nd ed. New York: McGraw-Hill Book Co. 1988.
11. Reijneveld SA. The choice of a statistic for testing hypotheses regarding seasonality. Am J Phys Anthropol 1990;83:181-184.

2.1.2. Studying seasonality by using a cosine function in regression analysis

A.M. Stolwijk, H. Straatman, G.A. Zielhuis

A statistical test which allows for adjustment of confounding can be required for the study of seasonal patterns. One way to realize this while retaining the information on the connection of time periods is by describing the seasonal pattern as a cosine function with variable amplitude and shift. After transformation of such a cosine function, it can be included into a regression model. The aim of this article is to supply for a detailed description of this method. An example of its application is given.

(submitted)

Introduction

Many studies have been published which concerned seasonal variation, for instance in births,[1] early pregnancy loss,[2] and in congenital malformations.[3] Whether a seasonal pattern exists can be studied in several ways. In this article we demonstrate a method that allows for adjustment of confounding. The first part is a general approach of studying seasonal patterns. In the second part we show a more detailed description of the method by means of mathematical functions. Subsequently, we give an application of this method using fictitious data. As an example we use the study of seasonality in fecundity, with ovulation as an indicator of fecundity.

Studying seasonality, a general approach

To study the seasonal variation in ovulation, data analysis can be performed in successive stages. The first step is to calculate and plot the prevalence of fecundity per month. Then confidence intervals surrounding the monthly prevalences can be calculated and added to the figure. From this information, it can be inferred whether there are differences in fecundity per month and whether there are differences between the months. In the same way, clusters of months can be formed and compared to other clusters of months. If confounding is presumed to occur, the next step is to adjust for such confounding effects. One way to perform this is by means of stratification so that insight can be gained into whether prevalences differ between months or clusters of months after adjustment for confounding. In this phase, problems may occur if several confounding factors are present. Adjustment for their effects simultaneously by means of stratification will often lead to small numbers of observations per month and thus to imprecise estimations of the prevalences. Nonetheless, these preliminary phases of analysis will provide the first indications of whether there is a specific seasonal pattern in fecundity. Rough evidence of such a pattern warrants a statistical test. In addition, a method is necessary which allows for adjustment of the effects of several confounders simultaneously.

We focus on the question of whether there is a seasonal pattern in fecundity during the course of a year, without paying attention to changes between years. To test for seasonality, a χ^2 test can be used to detect any departure from a uniform distribution. A more specific test should take into account the connection between time periods such as months or weeks. The method of Edwards[4] tests whether frequencies follow a sine function over 12 months. Also adaptations of the Edwards' test are suitable, for instance the one of Cave and Freedman[5] to test a bimodal seasonal pattern over 12 months, of Walter and Elwood[6] which can be used in the case of unequal populations at risk, of Roger[7] for small sample sizes and of Jones *et al*[8] for an arbitrary shape of the seasonal

effect. The nonparametric Hewitt's test[9] or its adaptation for other than 6-months periods by Rogerson[10] can also be applied, but they are less powerful than parametric tests. A Kolmogorov-Smirnov type statistic of Freedman[11] has a better power than the χ^2 test and the Hewitt's test in samples of moderate size. None of these tests allows for adjustment of confounding effects, except the method of Jones *et al.*[8] Moreover, some of them, including the latter, require special software. Therefore another test which allows for adjustment of confounding and that can be performed by widely available statistical computer programs is warranted.

In epidemiological practice, multivariate analysis techniques are commonly used to adjust for confounding. Linear regression analysis is often performed if the dependent variable has a normal distribution. In studies on seasonality in fecundity, the dependent variable is likely to be dichotomous, for example an ovulation is either present or it is not. In such a case, logistic regression analysis can be used.

To test whether fecundity is seasonally distributed, a cosine function with variable amplitude and shift can be introduced into the regression model. Depending on the hypothesis being tested, a cosine function with a period of one year, half a year or shorter can be included into the model. The maximum likelihood method estimates the regression coefficients for the best fitting regression line. The amplitude and the amount of shift of the cosine function can be calculated from the regression coefficients. Per time period, the probability of fecundity and the odds ratios can be calculated using the logistic regression model.

Detailed description of the method

A linear regression model can be developed to analyze seasonality in fecundity. Generally speaking, such a model will have the following form:

$$y = \beta_0 + \beta_{season} \cdot season + \beta_{C_1} \cdot C_1 + \ldots + \beta_{C_N} \cdot C_N$$

where β_0 is the intercept and C indicates a confounder; y is a particular parameter for fecundity and is normally distributed or can be transformed into such a distribution. An example of such a variable is the level of follicle stimulating hormone. In many studies on fecundity, the outcome parameter is defined as the probability of fecundity. This probability can be modelled in a logistic regression model such as:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_{season} \cdot season + \beta_{C_1} \cdot C_1 + \ldots + \beta_{C_N} \cdot C_N$$

where P is the probability of fecundity, for instance the probability of an ovulatory cycle. To define the variable 'season' in these models, it is hypothesized that the seasonal pattern

under study to follows a cosine function with variable amplitude and horizontal shift. In this cosine function, two periods must be defined: (i) the time period which defines the measure of fecundity, e.g. 'month' in 'the probability of an ovulatory cycle per month' and (ii) the period described by one cosine function. As an example we take 'month' as the time period under study, and 'one year' as the period of the cosine function. The cosine function can be described as:

$$f(x) = \alpha \cdot \cos(x - \theta) \tag{1}$$

where:

$$x = \frac{\pi}{t_n} + \frac{2\pi \cdot (t_i - 1)}{t_n} \quad (\textit{in radials}) \tag{2}$$

t_n = number of time periods described by one cosine function over [0, 2π) (e.g. t_n = 12 months in Figure 1);

t_i = ith time period (e.g. in Figure 1, for January: $t_i = 1$, for February: $t_i = 2$, etc.);

α = amplitude, >0;

θ = horizontal shift of the cosine function (in radials).

Thus, the logistic regression model will be as follows:

$$\ln(\frac{P}{1-P}) = \beta_0 + \alpha \cdot \cos(x - \theta) + \beta_{C_1} \cdot C_1 + \ldots + \beta_{C_N} \cdot C_N$$

As θ is unknown, transformation of this cosine function is required before the regression analysis can be performed. A rule of goniometry is used for the transformation of formula (1) into the next function which can be used in a regression model:

$$f(x) = \beta_1 \cdot \sin(x) + \beta_2 \cdot \cos(x) \tag{3}$$

This changes the logistic regression model into:

$$\ln(\frac{P}{1-P}) = \beta_0 + \beta_1 \cdot \sin(x) + \beta_2 \cdot \cos(x) + \beta_{C_1} \cdot C_1 + \ldots\ldots + \beta_{C_N} \cdot C_N$$

With x as given in formula (2), this leads to:

$$\ln(\frac{P}{1-P}) =$$

$$\beta_0 + \beta_1 \cdot \sin[\frac{\pi}{t_n} + \frac{2\pi \cdot (t_i - 1)}{t_n}] + \beta_2 \cdot \cos[\frac{\pi}{t_n} + \frac{2\pi \cdot (t_i - 1)}{t_n}] + \beta_{C_1} \cdot C_1 + \ldots + \beta_{C_N} \cdot C_N$$

Then the probability of fecundity in each time period can be calculated by:

$$P = \frac{e^{\beta_0 + \beta_1 \cdot \sin[\frac{\pi}{t_n} + \frac{2\pi \cdot (t_i - 1)}{t_n}] + \beta_2 \cdot \cos[\frac{\pi}{t_n} + \frac{2\pi \cdot (t_i - 1)}{t_n}] + \beta_{C_1} \cdot C_1 + \ldots + \beta_{C_N} \cdot C_N}}{1 + e^{\beta_0 + \beta_1 \cdot \sin[\frac{\pi}{t_n} + \frac{2\pi \cdot (t_i - 1)}{t_n}] + \beta_2 \cdot \cos[\frac{\pi}{t_n} + \frac{2\pi \cdot (t_i - 1)}{t_n}] + \beta_{C_1} \cdot C_1 + \ldots + \beta_{C_N} \cdot C_N}}$$

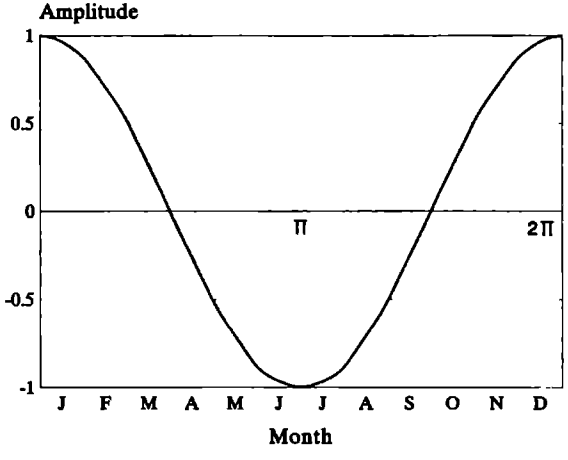


Figure 1. Twelve months described by one cosine function with amplitude = 1 and shift = 0 radials over [0, 2π)

Illustration of the transformation of formula (1) into formula (3) and presentation of the formulas to calculate the amplitude and the shift of the cosine function

The season described as formula (1) is the same as formula (3) for a given β_1 and β_2, which can be illustrated as follows.

Formula (3) can be rewritten:

$$\beta_1 \cdot \sin(x) + \beta_2 \cdot \cos(x) = \sqrt{\beta_1^2 + \beta_2^2} \cdot [\frac{\beta_1}{\sqrt{\beta_1^2 + \beta_2^2}} \cdot \sin(x) + \frac{\beta_2}{\sqrt{\beta_1^2 + \beta_2^2}} \cdot \cos(x)]$$

provided that β_1 and β_2 are not both zero.

Define:

$$\alpha = \sqrt{\beta_1^2 + \beta_2^2} \qquad (4)$$

$$\gamma_1 = \frac{\beta_1}{\sqrt{\beta_1^2 + \beta_2^2}} \tag{5}$$

$$\gamma_2 = \frac{\beta_2}{\sqrt{\beta_1^2 + \beta_2^2}} \tag{6}$$

Then formula (3) can be rewritten as:

$$\alpha \cdot [\gamma_1 \cdot \sin(x) + \gamma_2 \cdot \cos(x)]$$

As $-1 \leq \gamma_1 \leq 1$ and $-1 \leq \gamma_2 \leq 1$ and $\gamma_1^2 + \gamma_2^2 = 1$, this provides a unique $\theta \in [0, 2\pi)$ (see Figure 2), this can be written as:

$$\alpha \cdot [\sin(\theta) \cdot \sin(x) + \cos(\theta) \cdot \cos(x)]$$

By using the following rule of goniometry:

$$\sin(\theta) \cdot \sin(x) + \cos(\theta) \cdot \cos(x) = \cos(x - \theta) \qquad \textit{for all } x$$

we have:

$$\beta_1 \cdot \sin(x) + \beta_2 \cdot \cos(x) = \alpha \cdot \cos(x - 0)$$

The value of θ can be derived from γ_1 and γ_2 as follows:

The equation $\sin(\theta) = \gamma_1$ has, in general, two solutions in $[0, 2\pi)$:
If $\gamma_1 \geq 0$, the solutions are:

$$\theta_1 = \arcsin(\gamma_1) \qquad \textit{and} \qquad \theta_1^* = \pi - \theta_1$$

If $\gamma_1 < 0$, the solutions are:

$$\theta_1 = 2\pi + \arcsin(\gamma_1) \qquad \textit{and} \qquad \theta_1^* = \pi - \arcsin(\gamma_1)$$

The equation $\cos(\theta) = \gamma_2$ has, in general, two solutions in $[0, 2\pi)$:

$$\theta_2 = \arccos(\gamma_2) \qquad \textit{and} \qquad \theta_2^* = 2\pi - \theta_2$$

The value of θ is the one element in the intersection of $\{\theta_1, \theta_1^*\}$ and $\{\theta_2, \theta_2^*\}$, i.e. $\theta = \{\theta_1, \theta_1^*\} \cap \{\theta_2, \theta_2^*\}$.

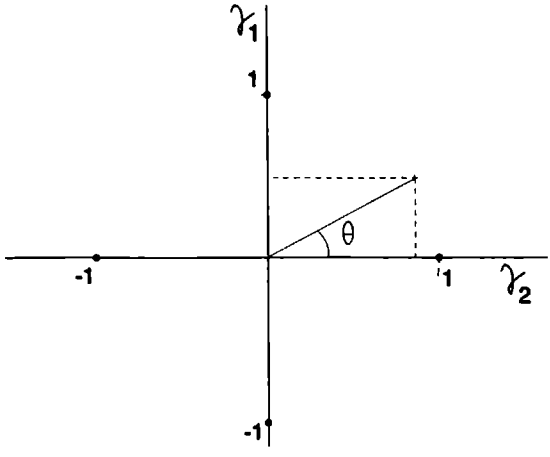


Figure 2. Unity circle, $\sin(\theta) = \gamma_1$, $\cos(\theta) = \gamma_2$

Application to fictitious data

To study seasonality in the occurrence of ovulation, data are gathered per month. The hypothesis to be tested is whether there is a seasonal pattern with one maximum level and one minimum level per year. Thus the time period is the 'month' and the period of the cosine function is 'one year'. Assume that the following seasonal pattern in the occurrence of ovulatory cycles was found by logistic regression analysis and that one confounder, age, is present:

$$\ln\left(\frac{P}{1-P}\right) = 1.15 + 1.02 \cdot \sin(x) + 0.34 \cdot \cos(x) - 0.03 \cdot age$$

Given the time period 'month' and a cosine function of 1 year, x [formula (2)] becomes:

$$x = \frac{\pi}{12} + \frac{\pi \cdot (t_i - 1)}{6}$$

Incorporation into the formula above results in:

$$\ln\left(\frac{P}{1-P}\right) =$$

$$1.15 + 1.02 \cdot \sin\left[\frac{\pi}{12} + \frac{\pi \cdot (t_i - 1)}{6}\right] + 0.34 \cdot \cos\left[\frac{\pi}{12} + \frac{\pi \cdot (t_i - 1)}{6}\right] - 0.03 \cdot age$$

Thus $\beta_1 = 1.02$ and $\beta_2 = 0.34$. To describe the seasonal pattern by the cosine function given in formula (1), the amplitude and shift have to be calculated. The amplitude is equal to [using formula (4)]:

$$\alpha = \sqrt{1.02^2 + 0.34^2} = 1.08$$

The shift can be calculated using formulas (5) and (6):

$$\gamma_1 = \frac{1.02}{\sqrt{1.02^2 + 0.34^2}} = 0.95$$

and

$$\gamma_2 = \frac{0.34}{\sqrt{1.02^2 + 0.34^2}} = 0.32$$

As $\gamma_1 \geq 0$, $\theta_1 = \arcsin(0.95) = 1.25$ radials and $\theta_1^* = \pi - \theta_1 = 1.89$ radials. $0_2 = \arccos(0.32) = 1.25$ radials and $\theta_2^* = 2\pi - \theta_2 = 5.03$ radials. The one element of the intersection is the shift, θ, which is 1.25 radials.
Thus the seasonal pattern can be described by:

$$1.08 \cdot \cos(x - 1.25)$$

The probability of ovulation can be calculated for every month:

$$P = \frac{e^{1.15 + 1.02 \cdot \sin[\frac{\pi}{12} + \frac{\pi \cdot (t_i - 1)}{6}] + 0.34 \cdot \cos[\frac{\pi}{12} + \frac{\pi \cdot (t_i - 1)}{6}] - 0.03 \cdot age}}{1 + e^{1.15 + 1.02 \cdot \sin[\frac{\pi}{12} + \frac{\pi \cdot (t_i - 1)}{6}] + 0.34 \cdot \cos[\frac{\pi}{12} + \frac{\pi \cdot (t_i - 1)}{6}] - 0.03 \cdot age}}$$

The results are shown in Figure 3. For each month, the probability of ovulation is calculated by assuming that the age of the women is 30 years. The probability of ovulation in March ($t_i = 3$) is 79% and in September ($t_i = 9$) 31%. If desired, the odds and adjusted odds ratios in each month can be calculated. The odds in each month is defined by [P/(1-P)], i.e. for March 3.76 and for September 0.44 in 30 year old women. The adjusted odds ratio in each month, independent of the effect of age, is calculated using the ratio of the odds in one month versus the odds in one reference month. For instance, the odds ratio for March is 8.55 if September is taken as the reference month. The amount of variation in the occurrence of ovulation explained by the season is expressed in the likelihood ratio test result [$-2\ln(L_1/L_2)$], which is the likelihood value from the model without the cosine function minus the likelihood value from the model with the cosine function. This value

can be compared with to a χ^2 distribution with 2 degrees of freedom to obtain a p value. A small p value indicates that the season makes a statistically significant contribution to the variation in fecundity. Whether the season has a biologically significant influence must be inferred from the prevalences.

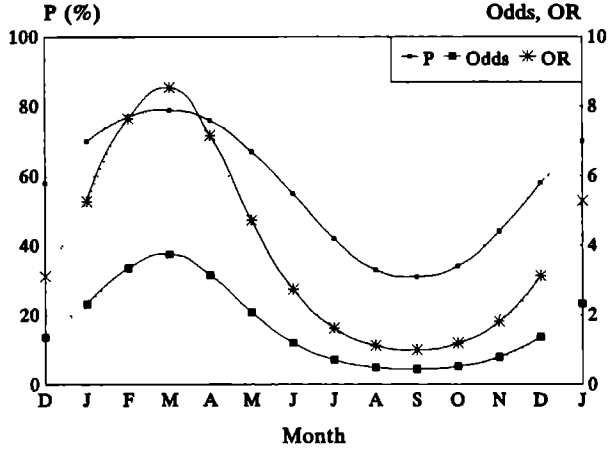


Figure 3. Probability of ovulation (P), odds and odds ratio (OR, with September as reference month) per month

Final remarks

One cosine function with variable amplitude and shift is a way of describing a seasonal pattern. However, a seasonal pattern might deviate greatly from such a cosine function. Then an analysis that uses one cosine function to test for seasonality may erroneously leads to the conclusion that there is no seasonal pattern. To determine whether another seasonal pattern is apparent, a mixture of cosine functions with different periods can be included into the model.

References

1. Matsuda S, Kahyo H. Geographical differences and time trends in the seasonality of birth in Japan. Int J Epidemiol 1994;23:107-118.

2. Weinberg CR, Moledor E, Baird DD, Wilcox AJ. Is there a seasonal pattern in risk of early pregnancy loss? Epidemiology 1994;5:484-489.

3. Bound JP, Harvey PW, Francis BJ. Seasonal prevalence of major congenital malformations in the Flyde of Lancashire 1957-1981. J Epidemiol Community Health 1989;43:330-342.

4. Edwards JH. The recognition and estimation of cyclic trends. Ann Hum Genet Lond 1961;25:83-87.
5. Cave DR, Freedman LS. Seasonal variations in the clinical presentation of Crohn's disease and ulcerative colitis. Int J Epidemiol 1975;4:317-320.
6. Walter SD, Elwood JM. A test for seasonality of events with a variable population at risk. Br J Prev Soc Med 1975;29:18-21.
7. Roger JH. A significance test for cyclic trends in incidence data. Biometrika 1977;64:152-155.
8. Jones RH, Ford PM, Hamman RF. Seasonality comparisons among groups using incidence data. Biometrics 1988;44:1131-1144.
9. Hewitt D, Milner J, Csima A, Pakula A. On Edwards' criterion of seasonality and a non-parametric alternative. Br J Prev Soc Med 1971;25:174-176.
10. Rogerson PA. A generalization of Hewitt's test for seasonality. Int J Epidemiol 1996;25:644-648.
11. Freedman LS. The use of a Kolmogorov-Smirnov type statistic in testing hypotheses about seasonal variation. J Epidemiol Community Health 1979;33:223-228.

2.1.3. Seasonality bias in poor reproductive outcome

A.M. Stolwijk, H. Straatman, G.A. Zielhuis

Epidemiology 1996;7:561-562 (letter)

Recently Basso *et al*[1] warned about bias in studies of seasonality in poor reproductive outcome, stemming from a seasonal variation in pregnancy planning. Such variation in pregnancy planning was found in a survey in five European countries during the 1970s and 1980s. For the same reason, bias can be expected in studies of seasonally-bound exposure and poor reproductive outcome. Examples of these forms of exposure are the use of pesticides and herbicides, the occurrence of influenza, exposure to sunlight, and the consumption of vitamin C. Bias will occur if (i) the exposure has a seasonal distribution, and (ii) the proportion of conceptions that end as a poor reproductive outcome varies throughout the season because of seasonal variation in pregnancy planning (as illustrated by Basso *et al*). The result can be a spurious relation between exposure and reproductive outcome, or, in the case of a real effect of exposure, over- or underestimation of the strength of the relation.

We studied the impact of this kind of bias in the association between exposure and poor reproductive outcome by means of a simulation study. Using the data presented by Basso *et al*, we defined three subpopulations, which differed in probability of conception and probability of spontaneous abortion (Table 1). We defined the distribution of the proportion of women who started per month as a cosine function with a period of 1 year and a shift of 7.5 months. This definition resulted in a seasonal pattern in pregnancy planning, with a maximum probability of 16.6% for starting in August to try to conceive and a minimum of 2.7% for starting in February. For simplicity, we used the moment of conception as the etiologic moment for spontaneous abortion. To maximize the overestimation, we defined the seasonal pattern in the probability of exposure as a cosine function with a period of 1 year and a shift of 3 months. We set the proportion of women exposed at 10%. These assumptions led to variation in the probability of exposure, with a maximum of 22.6% for conceptions in March and April and a minimum of 4.1% for conceptions in September and October. Note that the probability of exposure did not vary among the three subpopulations and thus was not related to the degree of fecundity, nor to the probability of spontaneous abortion. The simulation was based on 100,000 women per year who planned to become pregnant and continued for a period of 20 years. As the model was not stable during the first years of simulation, we analyzed only the results of the last 10 years.

The results are shown in Table 1. Model 1 used the data presented by Basso *et al* and resulted in a relative risk of spontaneous abortion for exposed vs. unexposed women of 1.03 [95% confidence interval (CI) = 1.01-1.05]. We found the largest bias when we defined a large fecund population (Subpopulations 1 and 2 combined) with extreme variation in the probabilities of spontaneous abortion (Model 3). This model resulted in

a relative risk of 1.10 (95% CI = 1.08-1.12). In all of the simulations, we found the highest probability of spontaneous abortions after conceptions in March or April and the lowest after conceptions in September.

We conclude that bias in the relation between a seasonally-bound exposure and poor reproductive outcome does occur because of seasonal variation in pregnancy planning, but, for practical purposes, this bias will be negligible.

Table 1. Simulations of seasonality in spontaneous abortion (SAB) risk due to exposure to X in a population with variation in distributions of fecundability and abortion risk (conditional on the null hypothesis of no real effect of X on abortion risk)

Model	Subpopulation						RR‡ (95% CI)	Range of SAB/ month (%)
	1: Fecund, low risk SAB		2: Intermediate		3: Subfecund, high risk SAB			
	P_C*	P_{SAB}†	P_C	P_{SAB}	P_C	P_{SAB}		
1. Basso *et al*[1]	36.7	6.3	21.7	8.8	9.2	14.5	1.03 (1.01-1.05)	9.5-10.8
2. More variation in P_{SAB}	36.7	3.1	21.7	8.8	9.2	29.1	1.06 (1.04-1.08)	12.5-16.5
3. Large fecund population	36.7	3.1	36.7	3.1	9.2	29.1	1.10 (1.08-1.12)	10.2-15.7

* P_C = probability of conception per month (in percentages).

† P_{SAB} = probability of spontaneous abortions per month (in percentages).

‡ RR = relative spontaneous abortion risk for exposure to X, relative to non-exposed.

Reference

1. Basso O, Olsen J, Bisanti L, Juul S, Boldsen J, and the European Study Group on Infertility and Subfecundity. Are seasonal preferences in pregnancy planning a source of bias in studies of seasonal variation in reproductive outcomes? Epidemiology 1995;6:520-524.

2.2. Is there seasonality in human ovulation?

A.M. Stolwijk, F.J.H. Aarts, C.J.C.M. Hamilton, P.H. Jongbloet, G.A. Zielhuis

To study seasonality in human ovulation in a direct way, we measured the occurrence of ovulation in infertile patients with spontaneous menstrual cycles (<6 weeks) who visited the fertility clinic at the University Hospital Nijmegen in the Netherlands for the first time in 1991 or 1992 (N = 407). Ovulation was detected using serial transvaginal ultrasound and midluteal progesterone measurement and was performed during one screening cycle. The frequency of ovulatory cycles per month varied from 73% to 93% (not statistically significant). No seasonal pattern in ovulation was found in subfecund Dutch women with spontaneous menstrual cycles. This finding was not confounded by the effects of age of the women, body mass index, or disorders that could influence ovulation.

Human Biology 1996;68:563-571

Introduction

Seasonality of births is found throughout the world.[1-3] Such a birth pattern can be influenced by social and cultural factors, but biological factors also may be involved.[2,3] For instance, decreased ovarian function was measured in women younger than 25 years and in women older than 35 years and in women with decreased energy intake or increased energy expenditure.[4] In several populations the changes in energy balance show seasonal variation.[5-7] Seasonal reproduction is common in mammals. For example, in rhesus monkeys seasonal variation in ovulation has been found.[8]

A mechanism that may cause seasonality in ovulation, other than influencing energy balance, is photoperiodicity. The information from the retina about light and darkness is transported by way of the suprachiasmatic nucleus in the hypothalamus to the pineal gland. The pineal gland produces melatonin from serotonin during darkness.[9] In humans, however, the role of the pineal gland and melatonin is still unclear.[10]

It is unknown whether birth seasonality in humans is caused by seasonal variation in ovulation. Indications of seasonal patterns of ovulation were found by Timonen *et al*[11] and Rameshkumar *et al.*[12] In Finland Timonen *et al* found less cystic glandular hyperplasia during the light season than during the dark season (30.1% vs. 34.6%) in 9750 endometrial biopsies. This finding indicated more ovulatory cycles during the light season (i.e., the months surrounding June). In contrast, an increase in anovulatory cycles was observed from May to July in the endometrial biopsies of 1036 women living in India.[12] In that study a negative correlation was found between the percentage of ovulatory cycles and the environmental temperature, whereas in Finland Timonen *et al*[11] found no relation between these two factors. Possibly, ovulation was suppressed because of low energy intake in India during spring and summer. Because these results are contradictory and based on indirect measurement of ovulation, we studied seasonality in human ovulation directly in a population with rather constant energy balance and moderate environmental temperature throughout the year. The convenient method of serial measurements of salivary progesterone cannot reliably discriminate between ovulatory and anovulatory cycles;[4] therefore we used the laborious method of serial transvaginal ultrasound and midluteal serum progesterone measurement.

Materials and Methods

In 1991 and 1992, 1021 couples visited the fertility clinic of the University Hospital Nijmegen, the Netherlands, for the first time. At the time of their first visit a detailed reproductive history was taken. The standard infertility workup consisted of a semen analysis, ultrasonographic ovulation detection during one cycle, assessment of cervical

mucus quality, and a timed postcoital test. The luteal phase was assessed by midluteal progesterone level seven days after follicular rupture and length of the luteal phase. The tubal status was determined by hysterosalpingography and/or laparoscopy. Hormonal screening was performed based on the menstrual history or if the history or clinical exam suggested an endocrine disturbance. Ultrasonographic ovulation detection was performed only in women with menstrual cycles shorter than 6 weeks (N = 422). Women with amenorrhea or oligomenorrhea (i.e., a menstrual cycle of 6 weeks or longer) were not screened because they were expected to have anovulatory cycles. In addition, women were not screened if they had been referred to the fertility clinic for specific treatments, such as in vitro fertilization or microsurgery. In this study retrospective data were used for the 422 women with spontaneous menstrual cycles less than 6 weeks.

The first transvaginal ultrasound scan was performed on cycle day 10 or, if the cycles were much shorter or longer than 28 days, 18 days before the expected onset of the next menstrual period. Scans were repeated on alternate days or daily, according to the follicular size. An ovulatory cycle was defined according to the ultrasonographic criteria of follicular rupture, that is, the observation of considerable loss of volume in a preexisting follicle, as described by Wetzels and Hoogland,[13] in combination with a midluteal progesterone level above 20 nmol/L. Luteinized unruptured follicle syndrome was diagnosed if the dominant follicle did not rupture but instead showed rapid growth after reaching a diameter of about 22 mm and remained a cystic structure throughout the luteal phase.[14]

Three groups of women at a higher risk for anovulatory cycles were distinguished on the basis of the following criteria: (i) age 35 years or older, (ii) body mass index (BMI) less than 20 kg/m^2 or more than 27 kg/m^2, and (iii) a disorder that could influence ovulation. BMI is defined as weight (kg)/height^2 (m^2). The cutoff point of 27 kg/m^2 was used because this is the criterion for obesity.[15] We use the general term 'disorders' to mean disorders that could influence ovulation, specifically, endocrine disorders (e.g., polycystic ovary syndrome, hyperprolactinemia), endometriosis, or cervical mucus disturbance. For the analyses concerning such disorders couples with unexplained infertility were excluded. Furthermore, a group of women at higher overall risk was formed from the women who met one or more of the three criteria.

To study seasonality, we used the month of the first day of the last menstrual period preceding the screening. Percentages of women with ovulatory cycles were computed for each month. Using logistic regression analysis, we studied whether these percentages differed between months and whether the seasonal variation followed a

pattern that could be described by a sine function with a unimodal or bimodal pattern. Therefore the months were entered into a logistic model as a sine function with a period of 0.5 year or 1 year and with variable amplitude and horizontal shift. Likelihood ratio test results gave information about the effects of including the season in the model on the explanation of the variation in ovulatory cycles. Furthermore, we studied whether the seasonal effect was more obvious in any of the groups of women at higher risk because of an interactive effect. Next, we performed logistic regression analysis for studying whether confounding by other risk factors influenced the relation between season and ovulation. Finally, logistic regression analysis was used to test whether the low percentage of women with ovulatory cycles found in a cluster of months differed from that in the rest of the year and if this difference remained after adjusting for confounding by other risk factors. We computed odds ratios with 95% confidence intervals for comparing the proportions of ovulatory cycles in the high- and low-risk groups.

To determine whether we could find a seasonal pattern with the available data, we calculated the power of the study. Because no standard formula for power calculation for seasonal variation is available, we used simulations for both a unimodal seasonal pattern (with the peak in June because of a possible influence of the photoperiod) and a bimodal seasonal pattern (with the peaks in June and December because of possible negative feedback mechanisms during the dark period).

Results

Four hundred twenty-two women were screened for ovulation. Most of the cycles were ovulatory (n = 354, 84%), but 53 (13%) were anovulatory (including 7 cycles with luteinized unruptured follicles). In 15 women the outcome of the screening was unknown because the screening cycle was incomplete or the ultrasonographic picture was ambiguous. These women were excluded from further analyses.

Characteristics of the remaining 407 women are given in Table 1. The youngest woman was 19 years old and the oldest was 45 years old; 15% of the women (n = 62) was 35 years or older. The BMI ranged from 16 kg/m^2 to 45 kg/m^2. Seventeen percent of the women (n = 60) had a BMI below 20 kg/m^2 and 10% (n = 35) a BMI above 27 kg/m^2. No data about height or weight were available for 55 women. Forty-three percent of the women (n = 132) had a disorder that could influence ovulation. In 98 couples the cause of infertility was unknown.

Table 1. Characteristics of the 407 women with ovulatory or anovulatory cycles

	Min.	Max.	Mean	SD	Median
Age (years)	19	45	30.5	4.3	31
BMI (kg/m^2)	16	45	22.8	3.9	22
Weight (kg)	42	126	64.1	11.7	62
Length (m)	1.48	1.89	1.68	0.07	1.68
	n		Unknown		%*
Primary infertility	272		0		67
Age <35 years	345		0		84
Age ≥35 years	62		0		15
BMI <20 kg/m^2	60		55		17
BMI >27 kg/m^2	35		55		10
Disorder†	132		98		43
Overall risk‡	227		93		72

* Percentage of the women with known values.

† Women with an endocrine disorder, endometriosis, or a cervical mucus disturbance (98 women with unexplained infertility were excluded).

‡ Defined as high if a woman met one or more of the following criteria: (i) 35 years of age or older, (ii) BMI less than 20 or greater than 27 kg/m^2, (iii) disorder.

In Table 2 the percentages of women with ovulatory cycles per month are given for the total population and for the groups of women at higher and lower risk; odds ratios are included. Although not statistically significant, the women aged 35 years or older had more ovulatory cycles than the women younger than 35 years. When 5-year age groups were distinguished, the group of 25-29-year-olds showed the fewest ovulatory cycles (82%) (Table 3). An ovulatory cycle was found in 78% of women with a BMI below 20 kg/m^2, in 90% of women with a BMI of 20-27 kg/m^2 and in 83% of women with a BMI above 27 kg/m^2 (Table 3). The women with a low or high BMI had fewer ovulatory cycles than those with an optimal BMI. Women with a disorder that could influence ovulation showed fewer ovulatory cycles than those without a disorder (Table 2). The group of women with one or more risk factors showed fewer ovulatory cycles than those with no risk factor (Table 2).

The percentages of ovulatory cycles throughout the year are shown in Figure 1. There were troughs in December-February, April, and September. The difference in the percentages of ovulatory cycles between months was not statistically significant. The season did not contribute to the explanation of the variation in ovulation: The pattern did not fit a model with the months included as 11 dummy variables or as a sine

Table 2. Women with ovulatory cycles per month in the total population, the groups at higher and lower risk for anovulatory cycle, and odds ratios

Month of screening	Total (N = 407)		Age (years) ≥35 (n = 62)		Age (years) <35 (n = 345)		Body mass index (kg/m²) <20, >27 (n = 95)		Body mass index (kg/m²) 20-27 (n = 257)		Disorder* Yes (n = 132)		Disorder* No (n = 177)		Overall risk† ≥ 1 risk factors (n = 227)		Overall risk† No risk factors (n = 87)	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
January	19	73	2	100	17	71	4	57	9	82	3	60	12	86	7	64	5	100
February	17	81	1	100	16	80	-	-	13	87	5	63	7	88	6	67	5	83
March	37	93	10	100	27	90	7	78	27	96	10	83	15	94	21	91	8	89
April	28	82	2	50	26	87	7	78	19	83	11	85	13	87	16	76	8	100
May	26	90	2	67	24	92	8	80	15	94	10	83	10	91	16	84	2	100
June	31	86	6	86	25	86	11	85	17	89	10	83	16	89	20	87	8	89
July	40	91	8	100	32	89	12	92	23	92	11	85	20	95	23	88	9	90
August	34	87	5	100	29	85	3	75	27	90	13	87	11	92	17	85	9	100
September	25	83	7	100	18	78	8	89	14	78	7	78	11	79	15	83	2	50
October	33	89	5	100	28	88	4	57	22	96	9	69	16	100	16	80	8	100
November	40	93	5	100	35	92	7	88	30	97	9	90	17	94	18	90	10	100
December	24	86	5	100	19	83	5	83	15	83	8	80	13	93	15	88	6	86
Total	354	87	58	94	296	86	76	80	231	90	106	80	161	91	190	84	80	92
OR (95% CI)‡			2.40 (0.83-6.89)				0.45 (0.24-0.86)				0.41 (0.21-0.79)				0.45 (0.19-1.05)			

* Women with an endocrine disorder, endometriosis, or a cervical mucus disturbance (98 women with unexplained infertility were excluded).

† Defined as high if a woman met one or more of the following criteria: (i) 35 years of age or older, (ii) BMI less than 20 or greater than 27 kg/m², (iii) disorder.

‡ Odds ratio with 95% confidence interval for the proportion of ovulatory cycles in the group at higher risk versus the group at lower risk.

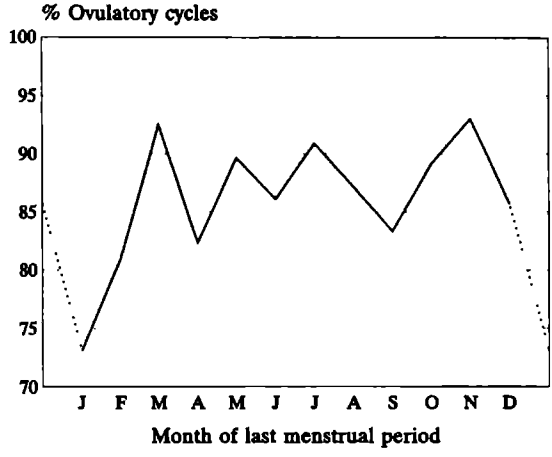


Figure 1. Percentage of women with ovulatory cycles per month (N = 407)

Table 3. Women with ovulatory cycles differentiated by age and BMI

	n	%
Age (years)		
19-24	26	90
25-29	106	82
30-34	164	88
35-39	49	94
40-45	9	90
Body mass index (kg/m^2)		
16-19	47	78
20-27	231	90
28-45	29	83

function with a period of 0.5 year or 1 year (p value of the likelihood ratio tests was always greater than or equal to 0.59). No interactive effects of the risk factors and season were found (Table 2). Moreover, the pattern was not masked by the effects of age, BMI, and disorders or by the effect of overall risk (p value of the likelihood ratio tests for the effect of season was always greater than or equal to 0.51). In addition, we tested whether the low percentage of women with ovulatory cycles during the period December-February (80%) was different from the percentage found in the rest of the year (89%). The crude analysis showed an odds ratio of 0.52 (95% confidence interval, 0.27-1.00), but after adjusting for the effects of age, BMI, and disorders, this difference

in ovulatory cycles disappeared considerably (odds ratio of 0.73; 95% confidence interval, 0.29-1.86).

If seasonal variation exists, with the given number of screenings per month we would have been able to detect a difference of about 10% in ovulatory cycles per month between the extremes of the sine function with $\alpha = 0.10$ and a power of 70%.

Discussion

This study did not reveal a seasonal pattern in ovulatory cycles of subfecund Dutch women in 1991-1992. The number of women in the study was rather small, resulting in a rather low power; therefore minor seasonal variation in ovulation might have been missed. The absence of a statistically significant seasonal variation in ovulation was not explained by confounding effects of age, BMI, and disorders.

The study population was highly selected. All the women had a history of infertility. The women with primary or secondary amenorrhea or with oligomenorrhea were not screened for ovulation and therefore were excluded from the study. The ovulation frequency in this study population does not represent the frequency in the general Dutch female population. However, seasonality itself is not likely to have influenced the selection process. Therefore, if seasonality in ovulatory frequency as a biological phenomenon does occur in the general population, such a pattern almost certainly would also appear in this selected study population, at least to a certain extent.

Ultrasonography in combination with serum progesterone measurement one week after follicle rupture is a reliable method for ovulation detection.[13] In 15 women we could not determine retrospectively whether or not ovulation had occurred, partly because of incomplete data in the patient dossiers. In January a relatively large number of screening results were inconclusive (4 out of 30 screenings), so the percentage of ovulatory cycles in that month would be at most 77%, that is, still one of the lowest percentages during the year. The inconclusive screening cycles change the results only slightly.

To study seasonality in ovulation, it is not necessary to follow women for several menstrual periods. In clinical practice (using the reliable method of serial transvaginal ultrasound in combination with midluteal serum progesterone measurements) following women for more than one period is even undesirable because the screening method is too cumbersome for the patients. Measurements during one menstrual cycle per woman give the same answer to the question so long as the women are equally distributed throughout the year for the known risk factors for anovulatory cycles and so long as

unknown risk factors are randomly distributed throughout the year. We presume that these assumptions are met because the fertility clinic did not change its procedure during the year. Moreover, we adjusted for possible confounding effects of age, BMI, and disorders. Thus this method was appropriate to study seasonality in ovulation.

The results of this study do not confirm the results of Timonen *et al*[11] in Finland or those of Rameshkumar *et al*[12] in India. The ovulation patterns found in those two studies were compatible with the results of Roenneberg and Aschoff,[3] who found a negative correlation between temperature and conceptions in equatorial regions with hot summers and a positive correlation in regions with cold winters and moderate summers. However, photoperiodicity and variation in environmental temperature are not the only features of seasonal influence; variation in energy intake and expenditure also matter. Therefore another explanation might be the decreased energy intake during spring and summer in India; Ellison[16] noticed decreased luteal function in that situation. No large seasonal variation in energy balance is expected in the Netherlands. Therefore in this study any influence of the season might be due to photoperiodicity and changes in environmental temperature. Because in the Netherlands (which is situated between 50° and 54° latitude) heterogeneity in photoperiodicity and temperature during the year is less than that in Finland, seasonal variation in ovulation may not be detectable.

In short, we conclude that no overt seasonality in human ovulation was present in these subfecund Dutch women with menstrual cycles shorter than 6 weeks.

References

1. Lam DA, Miron JA. The seasonality of births in human populations. Ann Arbor, Michigan: Population Studies Center. University of Michigan; 1987.

2. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: I. Biology, sociology or both? J Biol Rhythm 1990;5:195-216.

3. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: II. Environmental correlations. J Biol Rhythm 1990;5:217-239.

4. Ellison PT. Measurements of salivary progesterone. Ann N Y Acad Sci 1993;694:161-176.

5. Bailey RC, Jenike MR, Ellison PT, Bentley GR, Harrigan AM, Peacock NR. The ecology of birth seasonality among agriculturalists in central Africa. J Biosoc Sci 1992;24:393-412.

6. Panter-Brick C, Lotstein DS, Ellison PT. Seasonality of reproductive function and weight loss in rural Nepali women. Hum Reprod 1993;8:684-690.

7. Jasienka G, Ellison PT. Heavy workload impairs ovarian function in Polish peasant women. Am J Phys Anthropol 1993;Suppl. 16:117-118.

8. Walker ML, Wilson ME, Gordon TP. Endocrine control of the seasonal occurrence of ovulation in rhesus monkeys housed outdoors. Endocrinology 1984;114:1074-1081.

9. Tamarkin L, Baird CJ, Almeida OFX. Melatonin: A coordinating signal for mammalian reproduction? Science 1985;227:714-720.

10. Speroff L, Glass RH, Kase NG. Clinical gynecologic endocrinology and infertility. 5th ed. Baltimore, MD: Williams & Wilkins; 1994.
11. Timonen S, Franzas B, Wichmann K. Photosensibility of the human pituitary. Ann Chir Gynaec Fenn 1964;53:165-172.
12. Rameshkumar K, Thomas JA, Mohammed A. Atmospheric temperature & anovulation in south Indian women with primary infertility. Indian J Med Res 1992;96:27-28.
13. Wetzels LCG, Hoogland HJ. Relation between ultrasonographic evidence of ovulation and hormonal parameters: luteinizing hormone surge and initial progesterone rise. Fertil Steril 1982;37:336-341.
14. Hamilton CJCM, Wetzels LCG, Evers JHL, Hoogland HJ, Muijtjens A, De Haan J. Follicle growth curves and hormonal patterns in patients with the luteinized unruptured follicle syndrome. Fertil Steril 1985;43:541-548.
15. Must A, Dallal GE, Dietz WH. Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht2) and triceps skinfold thickness. Am J Clin Nutr 1991;53:839-846.
16. Ellison PT. Salivary steroids and natural variation in human ovarian function. Ann N Y Acad Sci 1994;709:287-298.

2.3. Seasonality in the results of in vitro fertilization

A.M. Stolwijk, M.J.C.M. Reuvers, C.J.C.M. Hamilton, P.H. Jongbloet, J.M.G. Hollanders, G.A. Zielhuis

Seasonal variation has been found in various reproductive outcomes. As known causes for reducing the rate of success of in vitro fertilization (IVF) cannot explain all the variation in IVF results, we studied whether the season had any additional explanatory power. The study population consisted of 1126 women who were treated for the first time with IVF at the University Hospital in Nijmegen, the Netherlands, between 1987-1993. Only first IVF cycles were analysed. After adjusting for confounding by the age of the woman, type of infertility, indication for IVF and year of aspiration, some seasonal variation was observed in the fertilization rate, embryo quality, pregnancy rate and birth rate.

Human Reproduction 1994;9:2300-2305

Introduction

Seasonal patterns have been found in various reproductive outcomes: in births throughout the world,[1-3] in stillbirths[4] and in spontaneous abortions.[4-6] These may be caused by seasonal variations in ovulation and sperm characteristics. Evidence to support such seasonality has been found for ovulation[7,8] and sperm characteristics.[9] It has been suggested that the season not only influences the number of ovulations but also the quality of the oocyte,[10] possibly by mediating gonadotrophins. Despite the fact that endogenous gonadotrophin secretion is suppressed by the hormonal therapy given during in vitro fertilization (IVF) treatment, it is still of interest to look for seasonality in the success rate of IVF.

Several factors are known to reduce the rate of success of IVF, e.g. advanced age of the woman,[11-14] the presence of primary infertility of the female,[14] the presence of male infertility[13,14] or multiple infertility factors.[13] However, these factors cannot explain the entire range of IVF results. The aim of this study was therefore to investigate whether season has any additional explanatory power. As the seasonal effect can be attributed to characteristics of the oocyte or spermatozoon, we studied the seasonal influence on the ovarian response to ovulation induction, the number of oocytes retrieved, fertilization rate, embryo quality, pregnancy rate and birth rate.

Materials and Methods

The study population consisted of all the women who were treated with IVF for the first time at the University Hospital Nijmegen, the Netherlands, between February 1987 and August 1993, for whom data had been recorded (N = 1154). In 1987 and the first half of 1988, data on only an aselect proportion of the women had been recorded. A group of 28 women were excluded from the analyses because their treatment cycle was cancelled for reasons other than poor response.

The season was expected to influence the oocyte during its development at the time of IVF treatment and the spermatozoon during its development before sperm collection on the day of oocyte aspiration. Therefore, to study seasonality, the date of aspiration was taken as an approximation of the moment of aetiological impact. In the treatment cycles which were cancelled because of low ovarian response to ovulation induction, the date of aspiration was approximated by adding 16 days to the first day of the last menstrual period because for the other women the median interval between the first day of the last menstrual period and aspiration was 16 days.

Successful IVF results were defined as: (i) sufficient ovarian response to ovulation induction, (ii) one or more fertilized oocytes, (iii) one or more transferred embryos of a

(fairly) good quality (i.e. greater than or equal to six cells on day 3, round embryo, ≤10% cytoplasmatic fragmentation and/or ≥90% equal blastomeres), (iv) pregnancy and (v) birth. Furthermore, differences in the number of oocytes retrieved were studied. Only the results of the first IVF treatment cycles were included in this study. Excluded were the results of cryopreserved embryos.

Potential confounding variables were considered to include the age of the woman, the type of infertility (primary or secondary) of the woman and indication for IVF, which consisted of a total of eight categories of combinations of tubal pathology, cervical factor or endometriosis and male infertility (while no donor spermatozoa were used) and unexplained infertility. Seven couples who were treated with IVF because of male infertility only and who used donor spermatozoa were excluded from the analyses which differentiated between the categories of indication. Furthermore, the year of aspiration was considered to be a potential confounding factor because the IVF treatment changed over the years [most important changes were the administration of gonadotrophin-releasing hormone (GnRH) agonist and the optimization of the time of aspiration].

The relationships between IVF results and the month of aspiration, age of the woman, type of infertility and indication for IVF were tested by χ^2 tests. Whether there were differences in the number of oocytes retrieved between the various categories was tested with the Wilcoxon-Mann-Whitney test or the Kruskall-Wallis one-way analysis of variance by ranks test. The association between the age of the woman and the number of oocytes retrieved was assessed by the Spearman correlation coefficient. The relationship between IVF results and the month of aspiration was corrected for the age of the woman, type of infertility, indication for IVF and year of aspiration by logistic regression. As a reference group the month of aspiration with the best IVF result was taken (provided that the number of IVF treatments in that month was relatively large). The results of the logistic regression analysis were expressed in odds ratios. The likelihood ratio test was used to examine whether a logistic regression model which included the month of aspiration and the confounding variables could explain more of the variation in the IVF results than a model with only the confounding variables.

Results

Almost two thirds of the women (712/1126) had a history of primary infertility. The age of the women varied from 20 to 47 years; half of the women were ≥33 years.

Table 1. Number and percentages of women with successful in vitro fertilization results per month of aspiration

Month of aspiration	Total	Response*		≥1 oocyte	≥1 fertilized oocyte†		≥1 'good' embryo†		Pregnancy†		Birth†	
	n	n	%	n	n	%	n	%	n	%	n	%
January	65	61	94	56	51	91	39	70	12	21	9	17
February	132	122	92	111	95	86	71	65	31	28	21	19
March	129	122	95	119	100	84	65	55	27	23	14	12
April	94	91	97	88	76	86	61	69	18	21	9	11
May	168	162	96	152	132	87	95	63	30	20	23	15
June	112	106	95	99	81	82	60	61	21	21	14	14
July	13	12	92	10	8	80	6	60	1	10	0	0
August	26	22	85	20	14	70	7	35	3	15	0	0
September	103	92	89	73	59	81	43	60	10	14	7	10
October	130	117	90	105	92	88	59	57	25	24	20	19
November	125	108	86	95	84	88	60	63	26	28	17	18
December	29	26	90	23	20	87	16	73	7	32	5	23
χ^2 df=11 (p)		19.89 (0.05)			9.38 (0.59)		14.78 (0.19)		10.03 (0.53)		14.25 (0.22)	
Total	1126	1041	92	951	812	85	582	62	211	22	139	15

* Calculated as part of the total number of women.

† Calculated as part of the number of women with one or more oocytes retrieved.

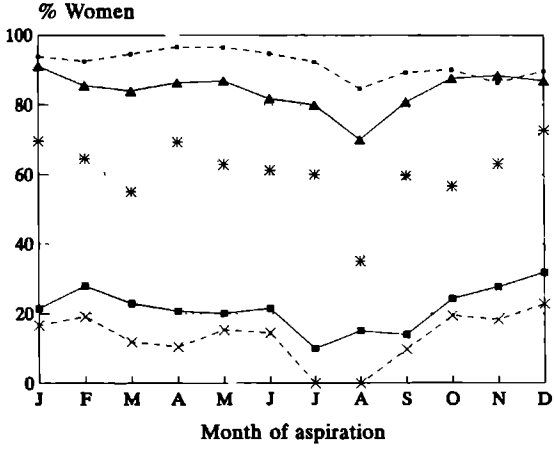


Figure 1. In vitro fertilization results per month of aspiration. Calculated as part of the total group of women (N = 1126): ▪ Response. Calculated as part of the women with ≥1 oocyte retrieved (n = 951): ▲ ≥1 fertilized oocyte, * ≥1 'good' embryo, ■ Pregnancy, x Birth.

Figure 1 and Table 1 present the response to the hormone treatment by month of aspiration. Of all the women treated, 8% did not respond to the hormone treatment. The poorest responses were observed during the second half of the year. As there were some differences in the response among women according to age group, primary or secondary infertility and category of indication (Table 2), and because the distribution of women in these categories over the year was unequal, logistic regression was performed to adjust for confounding. In addition the year of aspiration was considered to be a confounder: the mean response before 1990 was 87% whereas in 1990-1993 this was 95%. Furthermore, before 1990 56% of the IVF treatments were performed from July to December, whereas in 1990-1993 only 29% were performed in the second half of the year. After adjusting for confounding by age, type of infertility, indication for IVF and year of aspiration, no differences were found in the response between months (see Figure 3A).

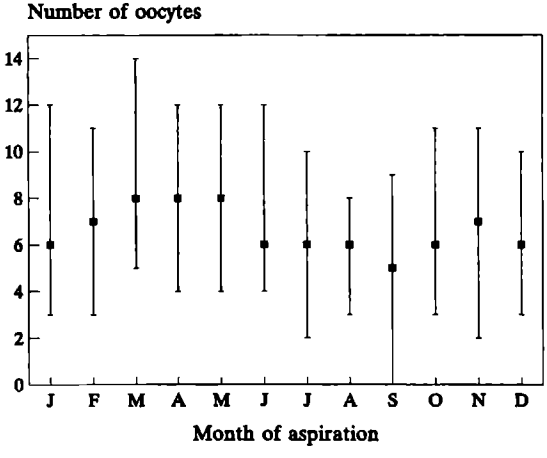


Figure 2. Number of oocytes retrieved per month of aspiration (25th, 50th and 75th percentile and Kruskall-Wallis χ^2, df = 11). I P25-P75; ■ P50; $\chi^2 = 32.69$ $p < 0.000$.

Information on the number of oocytes retrieved in each month is presented in Figure 2. In the group of 1126 women, the median number of oocytes was seven. No oocytes could be retrieved in 175 women: in 85 there was no ovarian response, in 84 aspiration was not performed because ovulation occurred after a luteinizing hormone (LH) peak and in six women aspiration was performed but no oocytes could be retrieved. The number of oocytes varied per month of aspiration (Figure 2), per year of aspiration (Kruskall-Wallis $\chi^2 = 103.93$, df = 6, $p < 0.001$) and in indication for IVF (Kruskall-Wallis $\chi^2 = 14.08$, df = 7, $p = 0.05$). The number of oocytes was negatively correlated with the age of the woman (Spearman $r = -0.15$, $p = 0.03$). The number of

oocytes in women with primary and secondary infertility did not differ (Wilcoxon Z = -0.76, p = 0.45). After adjusting for the age of the woman, type of infertility, indication for IVF and year of aspiration, the proportion of women from whom more than seven oocytes could be retrieved, i.e. above the median number, was higher in March than in September ($p < 0.05$; Figure 3B).

Differences in subsequent IVF results were studied conditional on the retrieval of one or more oocytes. Figure 1 and Table 1 show these results per month of aspiration. The proportion of women with one or more fertilized oocytes was the highest during April-May and October-February. In April-May and November-February the frequency of women with one or more embryos of (fairly) good quality was relatively high. The pregnancy rate was the highest for women with oocyte aspiration during February-March and October-December. The birth rate was the highest for women with oocyte aspiration in October-February. However, these differences in IVF results between months of aspiration were not statistically significant.

As could be expected, there were some differences in the results of IVF between the women with primary or secondary infertility, and between the categories of indication (Table 2). The women with secondary infertility had more success than the women with primary infertility. The indication for IVF seemed to influence the fertilization of the oocytes and the quality of the embryos. Couples who were treated with IVF exclusively because of male infertility had the least success. If one or more oocytes were retrieved, the age of the woman did not seem to have much influence on the subsequent IVF results (Table 2). Furthermore, in general, the IVF results up to 1990 were inferior to those in later years (data not shown).

To test whether there were monthly differences in the IVF results, it was considered appropriate to adjust for confounding in logistic regression models. On the condition that at least one oocyte was retrieved, the fertilization of one or more oocytes occurred more often in January than in August ($p < 0.05$; Figure 3C). The quality of the embryo was better in January than in March ($p < 0.05$; Figure 3D). The pregnancy rate was higher for women with oocyte aspiration in February than in September ($p < 0.05$; Figure 3E), and the birth rate was higher for women with oocyte aspiration in February than in April ($p < 0.05$; Figure 3F).

In testing whether the season was related to the response to hormonal treatment, number of oocytes retrieved, fertilization rate, quality of the embryo, pregnancy rate and birth rate for each of these IVF results, the logistic model with the month of aspirations and confounding variables was compared with the model with only the confounding variables. The log-likelihood ratio test was not statistically significant for

any of the IVF results, indicating that adding season to the model did not improve the explaining power of the variation in any of the IVF results.

Table 2. Number and percentages of women with successful in vitro fertilization results per age group, type of infertility and category of indication

	Total	Response*		≥1 oocyte	≥1 fertilized oocyte†		≥1 'good' embryo†		Preg-nancy†		Birth†	
	n	n	%	n	n	%	n	%	n	%	n	%
Age (years)												
20-24	14	14	100	13	11	85	8	62	3	23	2	15
25-29	194	187	96	164	137	84	98	60	36	22	21	13
30-34	558	523	94	482	418	87	302	63	111	23	75	16
35-39	313	273	87	251	211	84	148	59	54	22	37	15
≥40	47	44	94	41	35	85	26	63	7	17	4	10
χ^2 df=4		19.13			1.50		1.63		1.02		1.62	
(p)		(0.001)			(0.83)		(0.80)		(0.91)		(0.81)	
Type of infertility												
Primary	712	655	92	600	502	84	350	59	122	21	76	13
Secondary	414	386	93	351	310	88	232	66	89	26	63	18
χ^2 df=1		0.58			3.84		5.45		3.15		5.11	
(p)		(0.45)			(0.05)		(0.02)		(0.08)		(0.02)	
Indication ‡												
Tubal pathology	298	280	94	243	225	93	182	75	59	25	39	17
Cervix/endometriosis	96	88	92	82	73	89	53	65	17	21	11	13
Tubal+cervix/endom.	66	59	89	52	49	94	35	69	13	25	8	16
Male infertility	190	172	91	162	113	70	64	40	23	14	15	9
Male+tubal	155	143	92	130	110	85	75	58	32	25	21	16
Male+cervix/endom.	132	121	92	113	90	80	64	58	22	20	14	13
Male+tubal+cervix/endom.	55	53	96	49	43	88	32	67	12	25	9	19
Unexplained	109	103	95	98	91	93	63	64	22	22	15	15
χ^2 df=7		4.98			53.87		54.21		8.03		5.64	
(p)		(0.66)			(0.001)		(0.001)		(0.33)		(0.58)	
Total	1126	1041	92	951	812	85	582	62	211	22	139	15

* Calculated as part of the total number of women.

† Calculated as part of the number of women with one or more oocytes retrieved.

‡ Excluding 18 couples who were treated with IVF for other reasons and seven couples who were treated with IVF on the indication of male infertility only and who used donor spermatozoa.

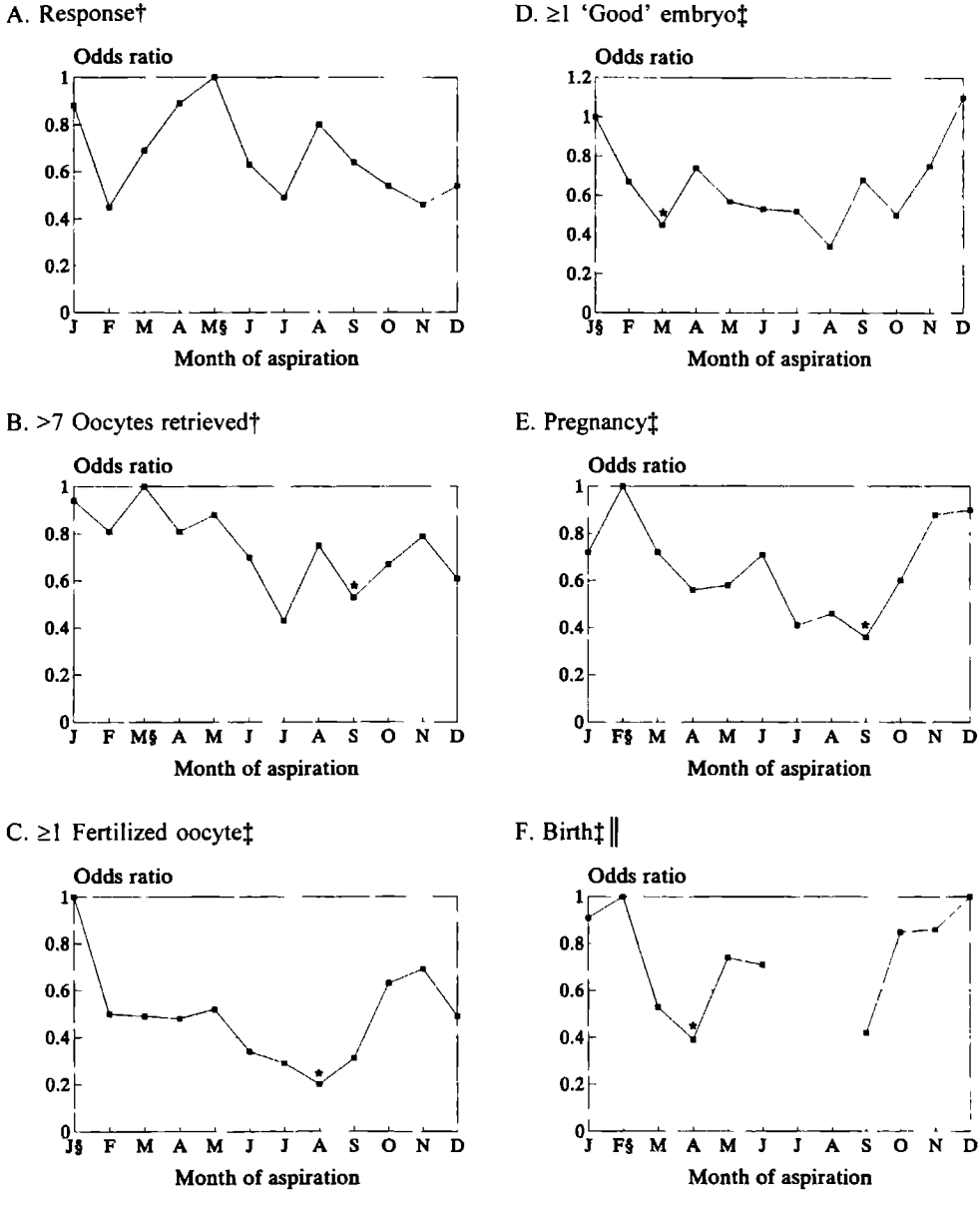


Figure 3. Odds ratios for each in vitro fertilization (IVF) result per month of aspiration, adjusted for age of the woman, type of infertility, indication for IVF and year of aspiration. § = Reference month (i.e. the month with the best results, provided that the number of IVF treatments in that month was relatively large). * = $p < 0.05$ as compared to the reference month. † = Calculated as part of the total group of women (N = 1126). ‡ = Calculated as part of the women with ≥1 oocyte retrieved (n = 951). ‖ = For July and August no odds ratio could be computed because there were no births.

Discussion

This study shows that seasonal differences in the results of IVF do occur, but some of the variation can be explained by age of the woman, primary or secondary infertility, indication for IVF and year of aspiration. This means that there is some random variation in the monthly distribution of the women by age, type of infertility and indication for IVF. In addition there is a year effect, as most of the relatively unsuccessful treatments before 1990 were performed in the second half of the year, whereas in 1990-1993 most of the treatments were carried out in the first half of the year and were more successful. Furthermore, as the number of IVF cycles per month was rather small, the estimates were not very precise. This was especially the case in July and August in which the IVF team had the summer recess. The tendency towards inferior results in these months may be due to imprecise estimates or to the recess itself. However, this was contradicted by the adjusted effects on the IVF results in the surrounding months, which also tended to be less successful than in the period November-February. This stability in results suggests a real effect of season with more favourable circumstances in the period November-February.

For the fertilization rate, the quality of the embryos and the pregnancy rate, it seemed appropriate, based on the pattern of Figure 3, to perform logistic regression models including the confounding variables and a cosine function of months [$\alpha \cos(x-\theta)$; $x = \pi/12 + \pi(m-1)/6$; where α = amplitude > 0; θ = horizontal shift of the cosine function]. From these models it appeared that the best results were found in the period November-February for the fertilization rate [log-likelihood ratio test $(-2\ln(L_1/L_2)) = 4.77$, df = 2, p = 0.09], for one or more 'good' embryo [$-2\ln(L_1/L_2) = 2.40$, df = 2, p = 0.30] and for the pregnancy rate [$-2\ln(L_1/L_2) = 4.52$, df = 2, p = 0.10]. As the numbers were rather small, it was not possible to perform these analyses on subsets, e.g. different periods, for looking at consistencies.

The biological mechanism through which the season may influence human reproduction is not clear. Melatonin, a hormone produced during darkness, may be one of the factors which influences gonadal function.[15] It should be emphasized that the hypothalamic-pituitary function of the woman is suppressed during the IVF procedure. It is therefore remarkable that there was still some seasonal difference in the number of oocytes retrieved which could not be explained by the age of the woman, type of infertility, indication for IVF and year of aspiration. Conditional on the retrieval of one or more oocytes, some seasonal differences in the subsequent results were still observed. This must have been related to either the inferior quality of the oocyte or the spermatozoon, or to diminished receptivity of the endometrium. A seasonal effect on

the spermatozoa may be most likely. In a review of seasonal variation in sperm characteristics, it was shown that the total spermatozoa per ejaculate and the sperm concentration were highest in the winter and spring and lowest in the summer, and the percentage of motile spermatozoa was highest in the spring.[9] Sperm characteristics may deteriorate through the influence of environmental temperature, which is highest during the summer months. As the production of spermatozoa takes ~74 days,[16] inferior spermatozoa may be found up till October. Analogous with this view, Paraskevaides *et al*,[17] showed that the conception rate was highest during October-March in women who underwent artificial insemination by donor. In that study maximum sperm counts occurred between February and May. However, no clear seasonal variation was found in the sperm characteristics of the partners (or the sperm donors) of the 951 women from whom one or more oocytes could be retrieved in our study.

In conclusion, seasonal variation was present, although some of the variation could be explained by the women's age, type of infertility, indication for IVF and year of aspiration. After correction, some monthly differences in the IVF results remained, with a tendency for better results following oocyte retrieval during the period November-February. As seasonal impact was not large, there seems no reason yet to take the season into account in IVF clinical practice.

References

1. Lam DA, Miron JA. The seasonality of births in human populations. Ann Arbor, Michigan: Population Studies Center. University of Michigan; 1987.

2. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: I. Biology, sociology or both? J Biol Rhythm 1990;5:195-216.

3. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: II. Environmental correlations. J Biol Rhythm 1990;5:217-239.

4. Sandahl B. A study of seasonal and secular trends in incidence of stillbirths and spontaneous abortions in Sweden. Acta Obstet Gynecol Scand 1974;53:251-257.

5. McDonald AD. Seasonal distribution of abortions. Br J Prev Soc Med 1971;25:222-224.

6. Kallan JE, Enneking EA. Seasonal patterns of spontaneous abortion. J Biosoc Sci 1992;24:71-75.

7. Timonen S, Franzas B, Wichmann K. Photosensibility of the human pituitary. Ann Chir Gynaec Fenn 1964;53:165-172.

8. Rameshkumar K, Thomas JA, Mohammed A. Atmospheric temperature & anovulation in south Indian women with primary infertility. Indian J Med Res 1992;96:27-28.

9. Levine RJ. Seasonal variation in human semen quality. In: Zorgniotti AW, (Eds.). Temperature and environmental effects on the testis. New York: Plenum Press, Seasonal variation in human semen quality. 1991;89-96.

10. Jongbloet PH. The effect of preovulatory overripeness of human eggs on development. In: Blandau RJ, (Ed.). Aging Gametes. Basel: S. Karger AG, The effect of preovulatory overripeness of human eggs on development. 1975;300-329.

11. Spira A. The decline of fecundity with age. Maturitas 1988;Suppl 1:15-22.
12. Padilla SL, Garcia JE. Effect of maternal age and number of in vitro fertilization procedures on pregnancy outcome. Fertil Steril 1989;52:270-273.
13. Tan SL, Royston P, Campbell S, *et al.* Cumulative conception and livebirth rates after in-vitro fertilisation. Lancet 1992;339:1390-1394.
14. FIVNAT. French national IVF registry: analysis of 1986 to 1990 data. Fertil Steril 1993;59:587-595.
15. Tamarkin L, Baird CJ, Almeida OFX. Melatonin: A coordinating signal for mammalian reproduction? Science 1985;227:714-720.
16. Speroff L, Glass RH, Kase NG. Clinical gynecologic endocrinology and infertility. 4th ed. Baltimore, MD: Williams & Wilkins; 1989:pp. 565-582.
17. Paraskevaides EC, Pennington GW, Naik S. Seasonal distribution in conceptions achieved by artificial insemination by donor. Br Med J 1988;297:1309-1310.

2.4. Seasonal variation in the time to pregnancy

2.4.1. Seasonal variation in the time to pregnancy: A secondary analysis of three Danish databases

A.M. Stolwijk, J. Olsen, I. Schaumburg, P.H. Jongbloet, G.A. Zielhuis

Three Danish databases were reanalysed to investigate seasonal variation in the time to pregnancy. Information was available on cohorts of women selected on the basis of union membership or residence in a given area: textile workers in Denmark (with 1,053 first and 1,771 second pregnancies), pharmacy assistants in Denmark (with 734 first and 725 second pregnancies) and pregnant women in the 36th week of pregnancy in two Danish cities (with 3,657 first and 3,526 second pregnancies). The influence of the season was of primary interest, because it is presumed to cause impaired ovarian function and hence a prolonged time to pregnancy. Furthermore, we studied whether the waiting time was prolonged in other situations with possibility of decreased ovarian function: in young and older women. In general, seasonality in the time to pregnancy based on the time of conception was found with a higher chance of a prolonged waiting time before conceiving in February-April and a lower chance of a prolonged waiting time before conceiving in August-October. This association was not distorted by the age of the women or diabetes mellitus. A prolonged time to pregnancy was found in women of 30 years or older. Women of 20 years or younger did not have a prolonged waiting time, but most of them were well beyond the age of menarche and thus beyond the period of impaired ovarian function. On a population level, there was evidence for seasonality in the time to pregnancy, which is compatible with seasonal variation in pregnancy planning as well as with biological influences.

European Journal of Epidemiology 1996;12:437-441

Introduction

The time necessary to achieve a clinically recognizable pregnancy varies considerably between couples. This time to pregnancy can be considered as a measure of fecundability.[1] The variation in the time to pregnancy may partly be explained in the literature by a variety of biological and environmental factors, such as: maternal age,[2] parity,[3] recent use of oral contraceptives,[4,5] smoking,[6,7] caffeine consumption[8] and occupational exposure to heat, noise, textile dyes, lead, mercury and cadmium.[9] In addition, median waiting times which preceded pregnancies ending in spontaneous abortions were 1.68 times longer than those in pregnancies resulting in live births.[10]

We hypothesize that impaired ovarian function is one underlying cause for a prolonged time to pregnancy. Impaired ovarian function is assumed to occur just after the menarche, before the menopause, after birth or abortion, after the use of oral contraceptives and in endocrine disorders such as diabetes mellitus. In these situations, more anovulatory cycles occur and consequently a prolonged time to pregnancy is expected. Anovulation in humans due to impaired ovarian function has also been hypothesized to vary with the season as a remnant of seasonal reproduction in mammals.[11] Indications for seasonality in human reproduction have been found for ovulation,[12,13] early pregnancy loss[14] and births.[15-17] The main theme of this study is whether seasonality in the time to pregnancy exists. Additionally, diminished ovarian function is studied by answering the question of whether in young and older women there is a prolonged time to pregnancy. We addressed these questions in a secondary analysis on three databases about specific Danish populations.

Methods

Three Danish databases were available, comprising information on female textile workers, pharmacy assistants and pregnant women in Aalborg and Odense. In this secondary analysis, only the information about the first and second pregnancies were used. Women who became pregnant despite contraception and women who reported fertility examinations or treatments were excluded from the study.

A. Textile workers

The first population consisted of all female textile, clothing and footwear workers in Denmark, who were members of a union in 1985 (N = 18,658).[10] By means of a postal questionnaire, these women were asked about all their pregnancies during 1979-1984. Response was 70.3%. For each pregnancy, information about the time to pregnancy was obtained by asking: 'For how long had you been trying to become pregnant before you succeeded? (Had regular intercourse without the use of any contraception)'.

Answer categories were: became pregnant despite contraception, 0-6 months, 7-12 months, 13-24 months, 25-36 months, more than 3 years and do not know. Questions about age at each pregnancy and about diabetes mellitus were included in the questionnaire. A total of 1,053 first and 1,771 second pregnancies met the criteria for analysis (Table 1).

B. Pharmacy assistants

The information in the second database was obtained in a study on 4,924 female pharmacy assistants in Denmark, who were members of a union at some time during 1979 to 1984.[18] The women were asked about all their pregnancies in the period 1979-1984 by means of a postal questionnaire; 92% responded. A total of 2,557 pregnancies were reported. As in the former paragraph, for each pregnancy, information was obtained about the time to pregnancy, age at conception and diabetes mellitus. No information was available about fertility examinations or treatments. The 734 first and 725 second pregnancies are shown in Table 1 according to age at conception, occurrence of diabetes mellitus and waiting time.

C. Pregnant women

The information in the third database was based on a questionnaire which was sent to all women in two cities in Denmark (Odense and Aalborg) who were in the 36th week of pregnancy between April 1984 and April 1987.[8] A total of 11,888 women participated in the study and completed the questionnaire (86%). Time to pregnancy was obtained by the question: 'For how many months had you and your partner been attempting to achieve conception; that is, how much time elapsed from when you stopped using contraception until you became pregnant?'. Answer categories were: 0-6 months, 7-12 months and 1 year or longer. Furthermore, information was available about diabetes mellitus. There were 3,657 first and 3,526 second pregnancies (Table 1).

For each database and for the first and second pregnancies separately, seasonality in the time to pregnancy was studied based on the time of conception. The seasonal influence would be the most obvious during the first menstrual cycles after the onset of trying to achieve a pregnancy, because long intervals will dilute the influence of the season. Given the information in the databases, we defined prolonged time to pregnancy therefore as a time to pregnancy of more than 6 months. Logistic regression models were created and likelihood ratio tests were performed to study whether adding the season to the model increased the explanation of the variance in the time to pregnancy. The season was included in the model as a cosine function with a period of one year and variable amplitude and horizontal shift. To adjust for confounding, age and diabetes mellitus were included in the model. Odds ratios per month of conception

Table 1. Distribution (percentages) of female textile workers, pharmacy assistants and pregnant women according to age, diabetes mellitus and time to pregnancy (TTP)

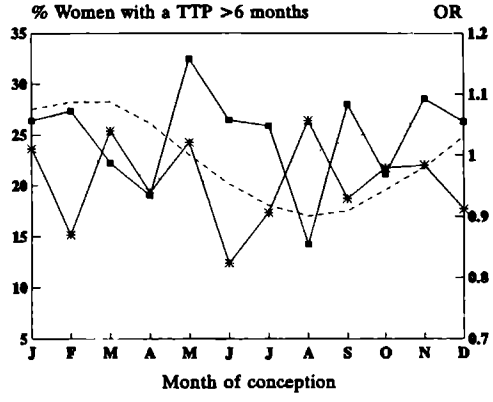
Characteristics	Textile workers		Pharmacy assistants		Pregnant women	
	1st pregnancy	2nd pregnancy	1st pregnancy	2nd pregnancy	1st pregnancy	2nd pregnancy
TTP >6 months	25.1	20.0	29.2	20.3	25.6	22.7
Age at conception						
≤20 years	36.4	19.0	37.8	15.2	16.9	5.2
21-29 years	59.7	72.8	60.1	78.5	75.6	74.3
≥30 years	4.0	8.2	2.1	6.4	7.5	20.6
Diabetes mellitus	0.3	0.4	0.4	0.6	0.4	0.4
Total (n)	1053	1771	734	725	3657	3526

were calculated, expressing for each month the ratio of the number of women with time to pregnancy of more than 6 months to the number of women with a time to pregnancy ≤6 months, compared to that ratio in a reference month. As reference month was chosen the month with the highest proportion of women with a prolonged time to pregnancy. To study the effect of age on the time to pregnancy, women who were 20 years or younger or 30 years or older at the beginning of the waiting time were compared to women of 21-29 years old. This was expressed in the form of odds ratios.

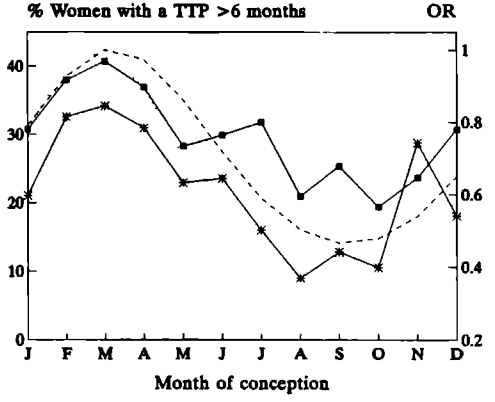
Results

Seasonal variation in the time to pregnancy was found in all three populations, both in the first and second pregnancies (Figure 1A-C). A relatively high proportion of the pharmacy assistants and the pregnant women had a long waiting time if they had conceived in February-April and a short waiting time if they had conceived in August-October. Almost the same pattern was observed in the textile workers as far as their first pregnancy was concerned. However, roughly the opposite pattern was found in the second pregnancies of the textile workers: a relatively high proportion had a long time to pregnancy if they had conceived in October-December and a short time to pregnancy if they had conceived in April-June. The seasonal variation in the textile workers was much smaller than in the other study populations; odds ratios varied from 0.9 to 1.1 in the textile workers, whereas in the pharmacy assistants and the pregnant women odds ratios varied from about 0.3 to 1.0 and 0.7 to 1.0, respectively. None of these patterns was confounded by the effects of age or diabetes mellitus.

A. Textile workers†



B. Pharmacy assistants‡



C. Pregnant women§

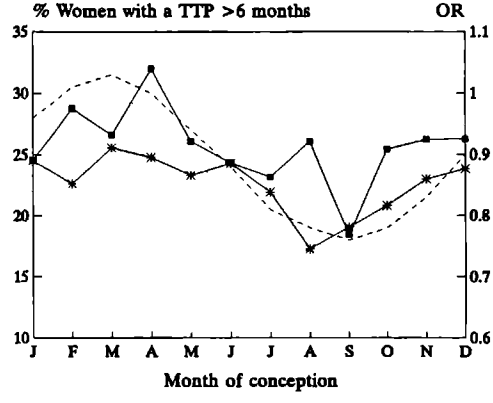


Figure 1.
Percentages of women with a time to pregnancy (TTP) of >6 months per month of conception, separated into first and second pregnancies. Odds ratios (OR) per month, expressing the ratio of the number of women with a time to pregnancy of >6 months to the number of women with a time to pregnancy of ≤6 months compared to that ratio in the reference month. The reference month is the month with the highest percentage of women with a time to pregnancy of >6 months. ■ % Women 1st pregnancy; * % Women 2nd pregnancy; ------- Odds ratio 1st pregnancy; ..... Odds ratio 2nd pregnancy.

† First pregnancies: unadjusted $-2\ln(L_1/L_2) = 0.84$ (df = 2, p = 0.66), adjusted for age and diabetes mellitus $-2\ln(L_1/L_2) = 21.90$ (df = 2, p = 0.000); second pregnancies: unadjusted $-2\ln(L_1/L_2) = 0.33$ (df = 2, p = 0.85), adjusted for age and diabetes mellitus $-2\ln(L_1/L_2) = 55.06$ (df = 2, p = 0.000).

‡ First pregnancies: unadjusted $-2\ln(L_1/L_2) = 11.34$ (df = 2, p = 0.003), adjusted for age and diabetes mellitus $-2\ln(L_1/L_2) = 10.75$ (df = 2, p = 0.005); second pregnancies: unadjusted $-2\ln(L_1/L_2) = 19.19$ (df = 2, p = 0.000), adjusted for age and diabetes mellitus $-2\ln(L_1/L_2) = 16.86$ (df = 2, p = 0.000).

§ First pregnancies: unadjusted $-2\ln(L_1/L_2) = 7.21$ (df = 2, p = 0.03), adjusted for age and diabetes mellitus $-2\ln(L_1/L_2) = 7.29$ (df = 2, p = 0.03); second pregnancies: unadjusted $-2\ln(L_1/L_2) = 7.38$ (df = 2, p = 0.03), adjusted for age and diabetes mellitus $-2\ln(L_1/L_2) = 6.95$ (df = 2, p = 0.03).

The odds ratios for the age categories are presented in Table 2. In the first pregnancies, the women of 20 years or younger conceived more quickly than the women aged 21-29 years. The results were less consistent in the second pregnancies: more young textile workers and pregnant women had a prolonged time to pregnancy than the women aged 21-29 years in the same populations. However, the confidence interval included unity. In each population, more of the women aged 30 years or older had a prolonged waiting time than those aged 21-29.

Table 2. Odds ratios (and 95% confidence intervals) for the women with a time to pregnancy of >6 months versus ≤6 months according to age

	Textile workers		Pharmacy assistants		Pregnant women	
	1st pregnancy	2nd pregnancy	1st pregnancy	2nd pregnancy	1st pregnancy	2nd pregnancy
Age at conception ≤20 versus 21-29	0.67 (0.49-0.91)	1.11 (0.82-1.50)	0.43 (0.30-0.62)	1.00 (0.59-1.69)	0.71 (0.57-0.88)	1.21 (0.85-1.74)
Age at conception ≥30 versus 21-29	1.74 (0.91-3.35)	1.79 (1.22-2.64)	1.61 (0.57-4.51)	3.16 (1.69-5.93)	1.58 (1.22-2.06)	1.81 (1.50-2.17)

Discussion

This analysis revealed a prolonged time to pregnancy in the women who conceived in February-April and a shorter time in those who conceived in August-October. This pattern was not found in the second pregnancy in the group of textile workers.

It should be noted that only women who had been or were pregnant were included in the three databases. Subfecund women with a longer time to pregnancy will have been under-represented in the study populations, especially in the group of pregnant women. As subfecund women have a higher probability of impaired ovarian function, selection might have led to underestimation of the effects of the season and age on the time to pregnancy.

The evidence for seasonal variation in the time to pregnancy found in this study may be the consequence of seasonal influences on the biological mechanism, for instance, a reaction of the pineal gland and melatonin to changes in the duration of light and darkness during the year,[19] or other seasonal changes in various types of exposure which are associated with low fecundity. Furthermore, the seasonal variation in the time to pregnancy may be the result of social factors. During the study period, the number of births in Denmark was relatively low during November-January and relatively high during March-July. As the use of contraceptives makes it possible to

plan pregnancies and as most women have a short time to pregnancy,[20] the birth pattern may be an indication for months that are less and more favoured for conceiving. This would lead to a relatively high percentage of women with a prolonged time before conceiving in February-April and a relatively low percentage of women with a prolonged time before conceiving in June-September. It is possible that this was applicable to our study. Thus, not only biological factors, but also a seasonal variation in pregnancy planning may have influenced the seasonal pattern in the time to pregnancy.

Unfortunately, the time to pregnancy was measured fairly roughly and recorded on the basis of the time of conception. Therefore we could not study seasonal variation in the time to pregnancy per month based on the precise starting date of the waiting period. A prospective study which includes all women at the start of their waiting time to pregnancy, whether they become pregnant or not, would give more insight into the biological role of seasonality in the time to pregnancy.

Waiting times were prolonged in the women older than 30 years. This is in accordance with our hypothesis of impaired ovarian function as cause for prolonged time to pregnancy and with the findings of others.[2] Results were not as clear for the youngest age group. In the first pregnancies, the youngest group of women conceived more quickly than the women aged 21-29 years; in the second pregnancies, there was some evidence of an opposite pattern, but this could be explained by chance, or it could be due to the selection of only pregnant women in the study sample. We had expected that a prolonged time to pregnancy would occur more often in young women, because after the menarche ovarian function is suboptimal. As most of the women aged ≤20 years fell outside this range, the fact that the young women did not have a prolonged time to pregnancy was probably due to the age classification itself. The number of women who conceived within three years after the menarche was too small to draw any conclusions. However, it should be noted that more of the second pregnancies in the women of 20 years or younger were preceded by still birth or spontaneous abortion than those in the older women (in the population of pregnant women in Aalborg and Odense: the ratio spontaneous abortions and still births versus live births in women ≤20 years was 0.70, in women 21-29 years 0.15 and in women ≥30 years 0.08). The occurrence of still births and spontaneous abortions may reflect decreased ovarian function in the youngest age group.

In conclusion, seasonality in the time to pregnancy based on the time of conception was found in the three databases. This is compatible with differential pregnancy planning as well as with biological influences.

References

1. Baird DD, Wilcox AJ, Weinberg CR. Use of time to pregnancy to study environmental exposures. Am J Epidemiol 1986;124:470-480.
2. Van Noord-Zaadstra BM, Looman CWN, Alsbach H, Habbema JDF, Te Velde ER, Karbaat J. Delaying childbearing: Effect of age on fecundity and outcome of pregnancy. Br Med J 1991;302:1361-1365.
3. Boldsen JL, Schaumburg I. Time to pregnancy - A model and its application. J Biosoc Sci 1990;22:255-262.
4. Linn S, Schoenbaum SC, Monson RR, Rosner B, Ryan KJ. Delay in conception for former 'pill' users. JAMA 1982;247:629-632.
5. Harlap S, Baras M. Conception-waits in fertile women after stopping oral contraceptives. Int J Fertil 1984;29:73-80.
6. Baird DD, Wilcox AJ. Cigarette smoking associated with delayed conception. JAMA 1985;253:2979-2983.
7. Kold Jensen T, Schaumburg I, Boldsen J. Cigaretrygning og ventetid til graviditet hos danske apoteksassistenter. Ugeskr Laeger 1992;154:1360-1363.
8. Olsen J. Cigarette smoking, tea and coffee drinking, and subfecundity. Am J Epidemiol 1991;133:734-739.
9. Rachootin P, Olsen J. The risk of infertility and delayed conception associated with exposures in the Danish workplace. Journal of Occupational Medicine 1983;25:394-402.
10. Schaumburg I, Boldsen JL. Waiting time to pregnancy and pregnancy outcome among Danish workers in the textile, clothing, and footwear industries. Scand J Soc Med 1992;20:110-114.
11. Jongbloet PH. The effect of preovulatory overripeness of human eggs on development. In: Blandau RJ, (Ed.). Aging Gametes. Basel: S. Karger AG, The effect of preovulatory overripeness of human eggs on development. 1975;300-329.
12. Timonen S, Franzas B, Wichmann K. Photosensibility of the human pituitary. Ann Chir Gynaec Fenn 1964;53:165-172.
13. Rameshkumar K, Thomas JA, Mohammed A. Atmospheric temperature & anovulation in south Indian women with primary infertility. Indian J Med Res 1992;96:27-28.
14. Weinberg CR, Moledor E, Baird DD, Wilcox AJ. Is there a seasonal pattern in risk of early pregnancy loss? Epidemiology 1994;5:484-489.
15. Lam DA, Miron JA. The seasonality of births in human populations. Ann Arbor, Michigan: Population Studies Center. University of Michigan; 1987.
16. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: I. Biology, sociology or both? J Biol Rhythm 1990;5:195-216.
17. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: II. Environmental correlations. J Biol Rhythm 1990;5:217-239.
18. Schaumburg I, Olsen J. Time to pregnancy among Danish pharmacy assistants. Scand J Work Environ Health 1989;15:222-226.
19. Speroff L, Glass RH, Kase NG. Clinical gynecologic endocrinology and infertility. 5th ed. Baltimore, MD: Williams & Wilkins; 1994:170-172.
20. Knuth UA, Mühlenstedt D. Kinderwunschdauer, kontrazeptives Verhalten und Rate vorausgegangener Infertilitätsbehandelungen. Geburtshilfe Frauenheilkd 1991;51:678-684.

2.4.2. Seasonal variation in the time to pregnancy: Avoiding bias by using the date of onset

A.M. Stolwijk, H. Straatman, G.A. Zielhuis, P.H. Jongbloet

To study seasonality in human fecundability, measured indirectly by time to the first pregnancy, we used data from 18,970 French-Canadian women who married for the first time during the 17th or 18th century. The time to pregnancy was approximated by the interval between marriage and first birth minus 38 weeks. We used the week of marriage and the week of conception as references to study seasonality. We found a minor seasonal pattern in time to pregnancy when using the week of marriage as a reference. The proportions of women with a short time to pregnancy were highest during December-January and June-July, indicating that these may be the most fecund periods. In contrast, we found an obvious seasonal pattern when using the date of conception as a reference. This pattern can be largely explained by a strong seasonal pattern in pregnancy planning (in this case, in marriages). When studying seasonal variation in the time to pregnancy, the date of onset of the time to pregnancy should be used as reference, not the date of conception. Otherwise, results will be biased owing to seasonality in pregnancy planning. The same is true for studies on seasonally bound exposures in relation to time to pregnancy.

Epidemiology 1996;7:156-160

Introduction

Seasonal reproduction is a common phenomenon in many animals. In humans, seasonal variation can also be observed for several reproductive factors, for instance, ovulation,[1,2] spermatozoa concentration,[3] pregnancy after artificial donor insemination,[4-7] early pregnancy loss,[8] and birth.[9-11] A seasonal pattern may be caused by photoperiodicity, which influences gonadal function. Information about the gonadal function of women can be derived from gonadal hormone production and ovulation. This information is difficult to gather, however. A more convenient, although indirect, measure of gonadal function is fecundability. The time to pregnancy can be considered as an indication of fecundability.[12]

For studying seasonality in the time to pregnancy, a date of reference should be defined. Either the date of the onset of the time to pregnancy (that is, the first day of refraining from any kind of contraceptive method because of pregnancy wish) or the date of the end of the time to pregnancy (that is, the date of conception) can be used. In this study, we used proxies for both definitions.

This article addresses the question of whether there is seasonal variation in the time to pregnancy. Data were obtained from a historical cohort of women who lived in Quebec during the 17th and 18th centuries. As contraceptive methods were not accessible in those days (with the exception of total abstinence from sexual intercourse), the date of marriage can be used as the date of the onset of the time to the first pregnancy. We approximated the end of the time to pregnancy by the date of the first birth minus 38 weeks for the gestational period. In addition, we studied the occurrence of bias related to the choice of the date of reference in studies on variation in the time to pregnancy.

Subjects and Methods

Family reconstitutions of the early French-Canadians are being compiled in the form of a population register called 'Le registre de la population de Québec ancien' by the Programme de recherche en démographie historique at the University of Montreal.[13] The register covers the entire population from the arrival of the first settlers in the early part of the seventeenth century to 1765, after the British takeover. This population can be considered to have lived in conditions of 'natural fertility,' that is, free from any contraceptive practice.[14]

In this study we used the interval between the date of marriage and the date of birth of the first child minus 38 weeks as a proxy for the time to pregnancy. We included women in the study if the dates of their marriage and first birth were known

and if the interval between their marriage and first birth was at least 240 days and, at most 1,096 days. We used the lower limit of 240 days because children born within a shorter interval had presumably been conceived before marriage. The upper limit was necessary because long intervals have often been found to correspond with a birth in the family that was missed by the registry. Because of the lower limit of 240 days, the time to pregnancy was below zero in some cases.

To study seasonality, we used both the week of marriage and the week of conception (approximated by the week of the first birth minus 38 weeks) as the date of reference. For reasons of simplicity, marriages or births on the 29th of February or on the 31st of December were excluded from the analyses based on the week of marriage or the week of conception, respectively (leading to exactly 52 weeks per year).

As seasonal influence would be the most obvious during the first menstrual cycles after marriage, because long intervals will dilute the influence of the season, we performed analyses to detect seasonal patterns in the time to pregnancy of ≤0, ≤1, ≤2, or ≤3 months. The weeks were entered into a logistic regression model as a sine function with a period of 26 weeks or 52 weeks with variable amplitude and shift. We chose the best fitting model as the one with the highest likelihood ratio. Additionally we calculated the deviance to determine whether the observed proportions of pregnant women per week were equal to the expected proportions when using the model with the best fitting sine function. We also used logistic regression analysis to correct for possible confounding effects of the woman's age at marriage (in three age groups) and the calendar year (in categories of 10 years and, in one case of small numbers, in a 20-year period).

Results

Data were available from 20,888 women. Only the first marriage was included in the analysis (N = 18,970). The age of the women at first marriage varied from 11 to 46 years. Half of the women married before the age of 21. Their marriages had taken place between 1634 and 1762.

Half of the women conceived within 3 months. The time to pregnancy varied with the age of the women: the youngest women (18 years or younger) and the oldest women (34 years or older) had the longest time to pregnancy. Moreover, the time to pregnancy tended to be the longest during the first decennia of French settlement in Canada.

The proportions of women with a time to pregnancy of ≤0, ≤1, ≤2, and ≤3 months per week of marriage and the best fitting sine functions are shown in Figure 1. A sine

function with a period of 26 weeks fitted better than one with a period of 52 weeks. There was only small variation in the proportions of women who became pregnant within 0, 1, 2, or 3 months: for ≤0 months from 4.5 to 6.4%, for ≤1 month from 22.9 to 27.3%, for ≤2 months from 38.9 to 45.0, and for ≤3 months from 51.6 to 54.8%. Because of the large number of observations, these seasonal patterns in the time to pregnancy deviated from a uniform distribution ($p < 0.002$ in most cases). Peaks in the bimodal curves were found during weeks 50-2 (December-January) and weeks 24-28 (June-July). The deviance (df = 49) was 76.97 for ≤0 months ($p = 0.01$), 69.09 for ≤1 month ($p = 0.03$), 58.93 for ≤2 months ($p = 0.16$), and 40.65 for ≤3 months ($p = 0.80$). These results indicated that the proportions of pregnant women per week of marriage could be predicted well with the sine function models with a period of 26 weeks. Adjustment for the effects of age at the time of marriage and the calendar period did not change the results to any substantial extent.

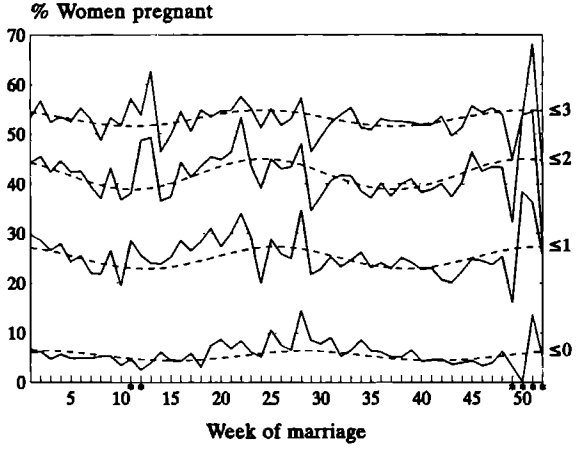


Figure 1. Percentages of women with a time to pregnancy of ≤0, ≤1, ≤2, or ≤3 months, per week of marriage; crude data and best fitting sine function (excluding marriages on the 29th of February or the 31st of December).
* = weeks with fewer than 50 marriages.

In contrast, when we used the week of conception as the date of reference, an obvious unimodal seasonal pattern was observed in the time to pregnancy (Figure 2). A sine function with a period of 52 weeks fitted better than a sine function with a period of 26 weeks. The deviance for the proportion of women with a time to pregnancy of ≤0, ≤1, ≤2, or ≤3 months was 312.12, 698.75, 248.92, and 106.63, respectively (df = 49, $p < 0.0001$). Thus, although the model with a sine function with a period of 52

weeks explained much of the variation in the distribution of the time to pregnancy throughout the year (likelihood ratio test result in all cases $p < 0.0001$), it did not predict the observed proportions very well. Moreover, the seasonal patterns in the time to pregnancy per week of conception remained after adjustment for confounding by age and calendar period.

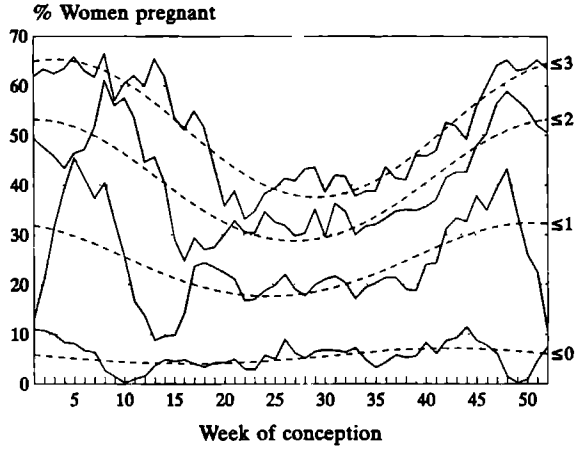


Figure 2. Percentages of women with a time to pregnancy of ≤0, ≤1, ≤2, or ≤3 months, per week of conception; crude data and best fitting sine function (excluding births on the 29th of February or the 31st of December)

Thus, the results of the two approaches for studying seasonality in the time to pregnancy differed considerably. The reason for this can be found in a seasonal pattern in marriages (Figure 3). The majority of women married during weeks 1-8 and 41-48 (that is, January-February and October-November). Because half of the women conceived within 3 months after marriage, a peak of conceptions followed during weeks 45-10 (November-mid-March) (Figure 4). The expected number of conceptions per week with a time to pregnancy of ≤0, ≤1, ≤2, or ≤3 months could be estimated by using the number of marriages per week (Figure 3) and the distribution of the time to pregnancy in the population. It appeared that the high proportion of women with a relatively short time to pregnancy after having conceived during the weeks 45-10 (Figure 2) was explained to a large extent by the distribution of marriages during the year. [For the four curves shown in Figure 2, the model χ^2 (df = 51) decreased from values of higher than 355, when assuming a uniform distribution, to values of 99 or less, when taking into account the distribution of marriages (Figure 3) and the time to pregnancy.]

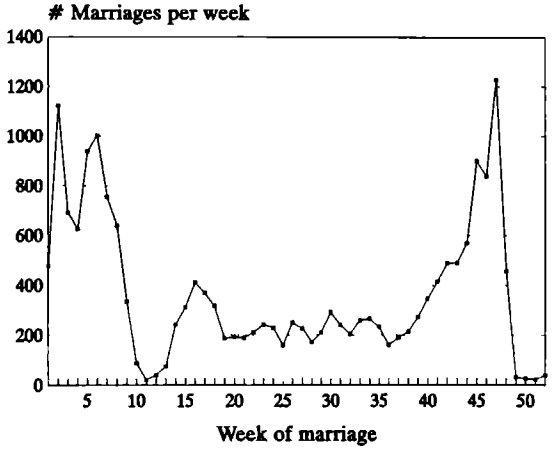


Figure 3. Number of marriages per week (excluding marriages on the 29th of February or the 31st of December)

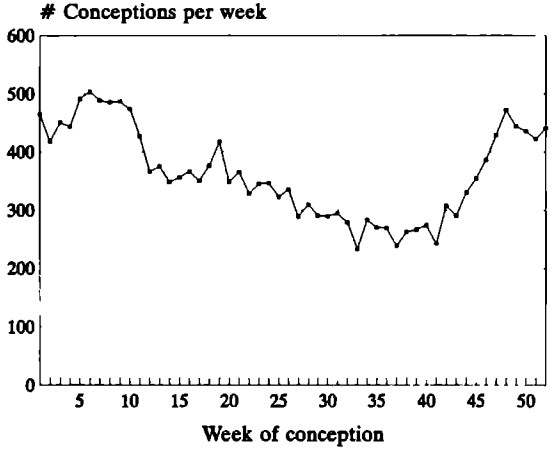


Figure 4. Number of conceptions per week (excluding births on the 29th of February or the 31st of December)

Discussion

As seasonal changes in, for instance, the photoperiod may cause a seasonal pattern in gonadal function and thus in fecundability, we focused on the question of whether there was a seasonal pattern in the time to pregnancy. The season was defined in two ways, that is, by the onset and by the end of the time to pregnancy. Although almost no seasonal pattern was observed in the time to pregnancy when the date of onset was

used as the point of reference, we found a striking seasonal pattern when the end of the time to pregnancy, that is, the week of conception, was the reference point. In this case, the seasonal variation in the time to pregnancy was explained by the seasonal pattern in marriages. Because 53% of the women became pregnant within 3 months, the marriage pattern almost dictated the conception pattern and thus the relation between the time to pregnancy and the date of conception.

We found the highest proportions of women with a short time to pregnancy among those who married in December-January or in June-July. This finding may indicate that gonadal function is optimal during these periods. With a part of the same database used here (N = 5,194), Nonaka *et al*[15] found that marriages in August-October resulted in a lower percentage of immediate conceptions (8-10 months after marriage) than in other seasons. Although this finding agrees with our results, Nonaka *et al* did not find a bimodal pattern. A possible explanation for the difference is that we used a database three times as large and studied seasonality in weeks instead of in 3-month periods. In Finland, Timonen *et al*[1] found the highest proportion of ovulations during the 'light season,' that is, around June. Rameshkumar *et al*,[2] in contrast, found the highest proportion of anovulations during the same period in India. These conflicting patterns, however, may be explained by the influence of both the photoperiod and temperature.[11] Spermatozoa concentration was found to be the highest during February and March and almost as high in November; other semen characteristics did not reveal a seasonal pattern.[3] In ovulating women who underwent artificial insemination by donor, the highest conception rates were found during October-March in England[4] and in October-January in Finland.[5] Unfortunately, the proportion of ovulating women per month were not reported in these two studies. In another Canadian study, Henderson-Toth *et al*[7] found the highest pregnancy rate after artificial insemination during November-December and February; they did not mention whether they excluded anovulatory cycles. Thus, although none of these other studies found a bimodal pattern, the indication in our study for a peak in fecundability during June-July was in agreement with the ovulation pattern found in Finland by Timonen *et al*,[1] whereas the peak in fecundability during December-January was in agreement with observations found in sperm concentration and after artificial insemination by donor.

A factor that possibly influences the seasonal pattern in the time to pregnancy might be seasonal variation in pregnancy loss. We lacked information to evaluate this issue. Several other studies found seasonality in spontaneous abortions when using births as a reference.[16-18] As seasonal variation in abortions can influence the number of births, the results of those studies may be biased. Two prospective studies provide

further insight. Weinberg *et al*[8] found peaks in the risk of early pregnancy loss (within 6 weeks after the last menstrual period) in 221 women with a positive pregnancy test result after having conceived in early September to early December. Nakamura *et al*[19] observed 11 spontaneous abortions in 519 women who had undergone ultrasonographic examination for confirmation of pregnancy. The last menstrual periods were more frequent during July-December (8/328) than during the first half of the year (3/189). The relatively high probability of pregnancy loss at the end of the year in these two studies seems to contradict the high proportion of women with a short time to pregnancy around December-January in our study. This discrepancy may indicate that the seasonal pattern in the time to pregnancy found in our study was not caused by a seasonal pattern in pregnancy loss.

To study a biological phenomenon that influences fecundability, it is correct to use the date of the onset of the time to pregnancy as a reference. In this case, the impact of a seasonal pattern in marriages or in pregnancy planning in general can be found only in the precision of the measured time to pregnancy per week of marriage; it cannot lead to bias. An important consequence of seasonality in pregnancy planning is the potential for biased results in studies on a seasonally bound exposure (for example, in an occupational setting) in relation to the time to pregnancy. This bias may occur if the exposure status is measured at the date of conception instead of at the onset of the time to pregnancy. This type of 'time bias' differs from the one discussed by Weinberg *et al*,[20,21] who related it to changes in exposure status over calendar time.

To overcome bias because of pregnancy planning, the onset of the time to pregnancy should be used as the date of reference. The best information would be obtained from a prospective study, in which women are enrolled in the study before they begin their time to pregnancy. This type of study, however, would have to cope with considerable practical problems connected with the large number of women who should be followed for a long period. An alternative is a retrospective study in which the date of the onset of the time to pregnancy is known. With such a study, the bias of pregnancy planning can be resolved. Unfortunately, if women are enrolled in the study while they are pregnant or after delivery, the date of the onset of the time to pregnancy and the date of conception will often be unknown and subfecund women will be underrepresented. The result will be underestimation of the strength of the relation, but no change in the direction of the effect estimators. In this study, the date of marriage was a reliable proxy for the onset of the time to pregnancy, because contraceptive methods were not accessible in the 17th and 18th centuries. Moreover, bias because of missing data of marriages or births is not likely, because the register is nearly

complete, and the population was essentially closed.[13]

We conclude that the impact of a biological factor in causing seasonal variation in the time to pregnancy and thus in fecundability, is small; the most fecund periods seem to be December-January and June-July. Moreover, to study this correctly, the onset of the time to pregnancy should be used as the date of reference, not the date of conception.

References

1. Timonen S, Franzas B, Wichmann K. Photosensibility of the human pituitary. Ann Chir Gynaec Fenn 1964;53:165-172.
2. Rameshkumar K, Thomas JA, Mohammed A. Atmospheric temperature & anovulation in south Indian women with primary infertility. Indian J Med Res 1992;96:27-28.
3. Levine RJ. Male factors contributing to the seasonality of human reproduction. Ann N Y Acad Sci 1994;709:29-45.
4. Paraskevaides EC, Pennington GW, Naik S. Seasonal distribution in conceptions achieved by artificial insemination by donor. Br Med J 1988;297:1309-1310.
5. Rönnberg L. Seasonal distribution in conceptions achieved by artificial insemination by donor. Br Med J 1989;298:187.
6. Woodworth SH, Pridham DD, Cook CL, Sanfilippo JS, Yussman MA. The effects of season on female fertility as measured by the outcome of frozen-thawed donor inseminations. Ann N Y Acad Sci 1994;709:199-200.
7. Henderson-Toth SM, Parker JA, Lommen CA, Martin JS, Toth DB, Yuzpe AA.Seasonal trends in pregnancy rates in a therapeutic donor insemination (TDI) program. Fertil Steril (Abstract). 1994;Suppl. 62:S182-S183.
8. Weinberg CR, Moledor E, Baird DD, Wilcox AJ. Is there a seasonal pattern in risk of early pregnancy loss? Epidemiology 1994;5:484-489.
9. Lam DA, Miron JA. The seasonality of births in human populations. Ann Arbor, Michigan: Population Studies Center. University of Michigan; 1987.
10. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: I. Biology, sociology or both? J Biol Rhythm 1990;5:195-216.
11. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: II. Environmental correlations. J Biol Rhythm 1990;5:217-239.
12. Baird DD, Wilcox AJ, Weinberg CR. Use of time to pregnancy to study environmental exposures. Am J Epidemiol 1986;124:470-480.
13. Légaré J. A population register for Canada under the French regime: context, scope, content and applications. Canadian Studies in Population 1988;15:1-16.
14. Charbonneau H. Les régimes de fécondité naturelle en Amérique du Nord: bilan et analyse des observations. In: Leridon H, Menken J, (Eds.). Natural fertility. Liège: Ordina Editions, 18, Les régimes de fécondité naturelle en Amérique du Nord: bilan et analyse des observations. 1979;441-491.
15. Nonaka K, Desjardins B, Légaré J, Charbonneau H, Miura T. Effects of maternal birth season on birth seasonality in the Canadian population during the seventeenth and eighteenth centuries. Hum Biol 1990;62:701-717.
16. McDonald AD. Seasonal distribution of abortions. Br J Prev Soc Med 1971;25:222-224.

17. Sandahl B. A study of seasonal and secular trends in incidence of stillbirths and spontaneous abortions in Sweden. Acta Obstet Gynecol Scand 1974;53:251-257.
18. Kallan JE, Enneking EA. Seasonal patterns of spontaneous abortion. J Biosoc Sci 1992;24:71-75.
19. Nakamura I, Uno M, Io Y, Ikeshita I, Nonaka K, Miura T. Seasonality in early loss of one fetus among twin pregnancies. Acta Genet Med Gemellol 1990;39:339-344.
20. Weinberg CR, Baird DD, Rowland AS. Pitfalls inherent in retrospective time-to-event studies: the example of time-to-pregnancy. Stat Med 1993;12:867-879.
21. Weinberg CR, Baird DD, Wilcox AJ. Sources of bias in studies of time to pregnancy. Stat Med 1994;13:671-681.

2.5. Seasonal variation in the prevalence of Down syndrome at birth: A review

A.M. Stolwijk, P.H. Jongbloet, G.A. Zielhuis, F.J.M. Gabreëls

Study objective: Many studies on seasonality in Down syndrome (DS) have been performed, leading to different results. It is hypothesized that seasonal variation in the hormone production by the hypothalamus-pituitary-ovarian axis just before ovulation leads to seasonality in conception rates of DS. The aim of this study is to determine whether there is seasonal variation in the prevalence of DS at birth as a proxy for seasonality in DS at conception. Design: We reviewed all the English and Dutch articles on this topic. Articles published between 1966 and January 1996 were traced by Medline, and by the reference lists. Main results: Twenty articles met the criteria for inclusion in this review. Although seven of these studies reported seasonality in DS prevalence, no consistent seasonal pattern was found in DS at birth in these seven studies, or in the other thirteen studies. A seasonal pattern could not have been masked by the effects of maternal age, induced abortions, shortened gestation or misclassification of DS. Conclusion: Seasonality in the prevalence of Down syndrome at birth does not exist. Thus, we found no support for the hypothesis that DS occurrence is related to seasonality in hormone production.

Journal of Epidemiology and Community Health (in press)

Introduction

It has been shown conclusively that the prevalence of DS at birth in any given population is related only to the maternal age of the childbearing population and to the use of prenatal diagnosis with subsequent termination of affected pregnancies.[1] The underlying mechanism for the higher prevalence of DS with advancing maternal age is not yet clear.

The most common hypothesis for the maternal age effect is ageing of the ovum itself.[2] Another hypothesis, the so-called compromised microcirculation hypothesis proposed by Gaulden,[3] states that hormonal imbalance causes a less-than-optimal microvasculature to develop around the maturing and mature follicles. This would result in an oxygen deficit and consequently in a decrease in the intra-cellular pH of the oocyte. The consequence would be a smaller size of the spindle, followed by displacement and nondisjunction of a chromosome. A more general hypothesis was proposed by Jongbloet[4] and presumes that hormonal imbalance may suppress the maturation of the oocyte during the follicular phase, which may be expressed, amongst others, by nondisjunctions. Both hypotheses about hormonal imbalance causing nondisjunctions may apply to seasonal influence.

It has been hypothesized that there is seasonal variation in human reproduction as most mammals show a seasonal pattern in reproduction. This pattern may be dictated by photoperiodicity which regulates the production of melatonin and inhibits or stimulates the production of gonadal hormones.[5] It has been hypothesized that remnants of such a seasonal reproduction pattern may still be present in humans and may cause seasonal variation in reproductive errors.[4] Seasonal variation in human reproduction has been observed in ovulations,[6,7] in sperm production,[8] in early pregnancy loss,[9] in spontaneous abortions[10-12] and in births.[13-15] As a consequence of seasonal variation in hormone production by the hypothalamus-pituitary-ovarian axis, a seasonal pattern in the prevalence of Down syndrome at birth can be expected.

For this article, we reviewed studies on seasonality in Down syndrome. As it is presumed that the etiologic moment occurs just before ovulation, our main interest was DS of maternal origin, preferably originating during first meiosis. Almost all DS cases (95 per cent) have free trisomy 21, which is a consequence of nondisjunction during meiosis one or two.[16] About 95 per cent of these extra chromosomes 21 are of maternal origin[17,18] and about 77 per cent of DS cases of maternal origin result from nondisjunction during the first meiosis.[19] Based on these arguments, seasonality in DS can be studied without making any distinction between the type of DS, parental origin or meiotic nondisjunction. As it is not possible to study the occurrence of Down

syndrome at conception, we reviewed articles about the occurrence of Down syndrome at birth. A consistent seasonal pattern in Down syndrome at birth will support our hypothesis that seasonality in Down syndrome occurs as a consequence of a seasonally-bound influence on the hormone production of the hypothalamus-pituitary-ovarian axis. However, if a seasonal pattern in Down syndrome is not obvious, this review cannot reject the hypothesis, as selective spontaneous abortions might have made such a pattern disappear.

Methods

A computerized literature search was performed by means of Medline. All the English and Dutch articles were selected which contained the words 'Down syndrome', 'Down's syndrome' or 'trisomy 21' in combination with 'season*' and were registered in the volumes published between 1966 and January 1996. Additionally, studies were traced via the reference lists of the articles. Unpublished work was not reviewed.

Studies were only included if they contained more than 50 DS cases, presented monthly results of DS prevalence at birth, and used a comparison group of total or live births in the corresponding period and area. Only original studies were included.

As photoperiodicity may influence hormone production and consequently the occurrence of nondisjunctions, the overview of seasonal patterns is arranged according to the latitude of the location of the study population. If a seasonal pattern exists, we expect to find a consistent seasonal pattern on the northern hemisphere and the opposite pattern on the southern hemisphere and that this pattern might be transient from the poles to the equator.

As in the original studies seasonality is analysed and interpreted in various ways, we used in this review the crude data of each study, i.e. the monthly DS rate compared to the average DS rate. As large differences existed in the overall DS prevalence between studies, presumably because of differences in maternal age, induced abortions after prenatal diagnosis and in registration, DS prevalence per month was not compared and clearly could not be pooled. Moreover, exact numbers were not always presented in the studies.

Results

In total, 53 English and Dutch publications were traced via Medline; only 13 of them met all the criteria for inclusion in this review. In Table 1 the main reason for exclusion is mentioned for each study. In addition, seven studies which met the criteria were found via reference lists.

Table 1 Main reason for excluding publications from this review

Reason	Publication
Not about seasonality in DS	
	Kessler & Lilienfeld (Advances in Cancer Research 1969,12 225-302)
	Chen & Woolley (Journal of Medical Genetics 1971,8 153-159)
	Kenna *et al* (Quarterly Journal Medicine 1975,44 17-44)
	Safra *et al* (Teratology 1976,14 143-149)
	Ikeda *et al* (Journal of Mental Deficiency Research 1977,21 139-151)
	Paradise (Pediatrics 1980,65 917-943)
	Klein *et al* (Journal of Pediatric Surgery 1984,19 370-374)
	Jongbloet *et al* (Diabetes Research 1988,9 51-58)
Not an original study	
	Pergament (The Chigaco Medical School Quarterly 1969,28 57-67)
	Jongbloet (The Lancet 1970,2 1317-1318)
	Lowe (British Medical Journal 1972,3 515-520)
	Hecht (In Hook & Porter (Eds) Population cytogenetics Studies in humans New York Academic Press, Inc, 1977, 237-250)
	Stark & White (In Hook & Porter (Eds) Population cytogenetics Studies in Humans New York Academic Press, Inc, 1977, 275-283)
	Ament (American Journal of Epidemiology 1976,103 342-343)
	Rothman (American Journal of Epidemiology 1976,104 585-586)
	Janerich & Jacobson (The Lancet 1977a,1 515-516)
	Janerich & Jacobson (The I ancet 1977b,1 1004-1005)
	Robinson (Advances in Pathobiology 1977,6 214-226)
	Robinson & Puck (The Lancet 1977,2 981-982)
	Sever (The Lancet 1977,1 754)
	Mikkelsen (Human Genetics 1981,2 (suppl) 211-226)
	Jongbloet (The Lancet 1983,2 347-348)
	Anonymus (The Lancet 1983,1 1312-1313)
	ICPEMC (Mutation Research 1986,175 263-266)
Other publication on (almost) the same population included in this review	Nielsen *et al* (Humangenetik 1973,19 67-74)
	Nielsen *et al* (Annales de Genetique 1981,24 212-215)
Less than 50 DS cases	
	Robinson & Puck (American Journal Human Genetics 1967,19 112-129)
	Haynes *et al* (Neurology 1974,24 691-700)
	Seifert & Sommer (American Journal of Diseases of Children 1986,140 822-824)
	Drugan *et al* (Fetal Therapy Clinical Advances 1989,4 195-199)
No DS prevalence per month	
	Leck (The Lancet 1966,2 457-460)
	Halevi (British Journal of Prevention and Social Medicine 1967,21 66-77)
	Baird & Miller (British Journal of Prevention and Social Medicine 1968,22 81-85)
	Hook *et al* (The Lancet 1974,1 566-567)
	Iselius & Lindsten (Human Genetics 1986,72 133-139)
No DS prevalence at birth	
	Jongbloet (Clinical Genetics 1971,2 315-330)
	Jongbloet (In Blandau (Ed) Aging Gametes Basel S Karger AG, 1975, 300-329)
	Puri & Singh (Br J Clin Pract 1995,49 129-130)
Not compared to total or live births in corresponding period and place	Jongbloet *et al* (Human Genetics 1982,62 134-138)
	Jongbloet & Vrieze (Human Genetics 1985,71 241-248)

In Table 2, an overview is given of the remaining 20 studies on seasonality in DS. Because of the influence of photoperiodicity on hormone production, the studies are listed by latitude. If two studies were performed at the same latitude, the one with the largest number of DS cases is mentioned first. The authors of 13 studies concluded that no relation was present between the month of birth and DS prevalence[20-32]. Seven studies[33-39] reported seasonality in DS prevalence.

These seven studies[33-39] were not a selective group in the number of DS cases, the prevalence of DS or latitude. Their numbers of DS varied from 103 to 2,469, and in the other 13 studies from 139 to 3,810. The prevalence of DS was in the range from of 0.88 to 2.4 per 1,000 in the seven studies, and from 0.43 to 1.86 per 1,000 in the other studies. The seven studies were located between Scotland and Victoria, Australia. Overall, no consistent pattern was found in seasonal variation in these seven studies that reported a seasonal pattern, or in the trends reported in the other studies. An unexplained cluster of relatively high prevalences of DS births might be apparent in November/December at the extreme end of the northern hemisphere,[21,24,26,27,29,36,38] but in the other months there was no consistent pattern. As the studies with a low number of DS cases may have missed a seasonal pattern, we focused on studies with more than 1000 DS births which gave information on the monthly prevalence of DS.[21,24,25,28,29,33,35,38,40] In this selective group of studies, no comparable seasonal pattern was found in DS at birth. They showed the same direction in prevalence of DS in comparison with the average prevalence per month only in two single months. There was a relatively high prevalence of DS births during August and a relatively low prevalence during June in the northern hemisphere, while the opposite pattern was observed in the southern hemisphere. During the other months, some of these nine studies reported a relatively high prevalence of DS, whereas the others showed a relatively low prevalence.

Discussion

In this review, no obvious seasonal pattern was found in DS prevalence at birth. Some factors can influence the DS prevalence at birth: maternal age, induced abortions after prenatal diagnosis[41] and shortened gestation.[42] Maternal age can only confound the relation between the season and the occurrence of Down syndrome if pregnancy planning during the year differs according to woman's age. As this is not apparent, confounding by maternal age is not expected. If there is a seasonal pattern in DS at conception, induced abortions after prenatal diagnosis might have weakened this pattern, but would not have changed the positions of the peaks and troughs of a

Table 2 Studies on seasonality and DS, listed by latitude

Study	Number of DS cases	Study population	Frequency per 1,000	Compared to births*	Conclusion in study about seasonality in DS	Higher and lower DS birth prevalence†											
						J	F	M	A	M	J	J	A	S	O	N	D
Northern hemisphere																	
Leisti *et al*[26]	263	live births in northern Finland 1965-1979	1 73	alive	no (NS‡)	-	+	-	-	+	+	+	-	-	-	+	+
Källén & Måsbäck[29]	1,174	born in Sweden 1973-1983	1 08	alive	no	-	+	-	+	-	-	-	+	+		+	
Lindsten *et al*[24]	1,370§	live births in Sweden 1968-1977	1 28	alive	no	-	-	-	+	+	-	-	+	+	-	+	+
Holloway & Emery[36]	978	born in Scotland, registered by 13 Health Boards, 1960-1974	0 88	alive	yes (NS), high in Jun, Nov, low in Jul	-	-	-	+	+	+	-		+		+	+
Videbech & Nielsen[38]	1,972	registered in Denmark (half of DS pop born before 1965), 1968-1980	unknown	alive	yes, high in Oct-Jan ($p < 0\ 05$), low in Feb-May, Apr-Sept	+	-	-			-	+	+	-	+	+	+
McDonald[21]	2 398	live and stillbirths in Quebec 1958-1967	1 86	alive	no	+	-	+	+	-	-	+	+	+	-	+	+
Baird & Sadovnick[27]	883 3	live births and 70% of selective aborted fetuses in British Colombia 1964-1983	1 21	alive	no	-	+	-	-	-	-	+	+	-	+	+	+
Leck[17]	527	born in Birmingham, United Kingdom, 1950-1965	1 64	total‖	yes, high in Jan-Jun	+	+	+		+	+	-	-	-	+	-	+
Knox & Lancashire[32]	354	live and stillbirths in Birmingham, United Kingdom, 1964-1984	1 38	total	no		+	+	-	+	-	+	-	+	-	-	-
Stoll *et al*[31]	139	live and stillbirths in Strasbourg and surrounding areas, France, 1979-1987	1 17	normal	no	¶											
Czeizel[28]	1 997	live births in Hungary 1970-1984	0 85	total	no (NS), high in Jul Dec, low in Jan, Jun	¶( )					( )	(+)					(-)
Stark & Mantel[40]	2,431	live births in lower peninsula of Michigan 1950-1964	0 89	alive	no (NS), high in Jun, low in Oct	-	-	+	+	-	+	+	-	-	-	+	-
Rothman & Fabia[35]	2 469	live births in Massachusetts 1950-1966	1 35	total	yes ($p = 0\ 03$), high in summer, low in winter	¶( )	( )	( )				(+)	(+)	(-)			

(Continued on the next page)

Table 2 Studies on seasonality and DS, listed by latitude (Continued)

Study	Number of DS cases	Study population	Frequency per 1 000	Compared to births*	Conclusion in study about seasonality in DS	Higher and lower DS birth prevalence†											
						J	F	M	A	M	J	J	A	S	O	N	D
Northern hemisphere																	
Castilla *et al*[30]	618	born in hospitals in Italy, 1981-1984	1 22	total	no (NS)	¶											
Wehrung & Hay[20]	3.810	live births in 29 states and 2 cities of the US 1962-1965	0 43	alive	no (NS)	¶											
Gummere *et al*[25]	1,364	born in Ohio 1970-1979	1 22	alive‖	no (NS) high in Jan, May, low in Apr, Jun	+	+	-	-	+	-	+	+	-	-	-	-
Kanai & Nakamura[39]	291	born in Kyoto, Japan, 1959-1979	unknown	total	yes	-	-	-	-	+	+	+	+	+	+	-	-
Kaplan & Ament[23]	299	born in Jerusalem, Israel, 1970-1972	1 03	total	no	+	+	+	-	-	-	+	-	-	+	+	+
Harlap[34]	103	born in Jerusalem, Israel, 1964-1970	2 4	alive	yes (6-month cycle $p < 0\,001$), high in spring, autumn	-	+	+	+	+	-	-	+	+	+	-	-
Southern hemisphere																	
Castilla *et al*[30]	919 (n-t), 348 (t)	born in hospitals in non-tropical (n-t) and tropical (t) South America 1982-1986	1 58 (n-t), 1 21 (t)	total	no (NS)	¶											
Collmann & Stoller[33]	1,134	live births in Victoria, Australia, 1942-1957	1 45	alive**	yes, low in Jul ($0\,01 < p < 0\,02$)	+	+	+	+		+	-	-	-	+	+	-

* From the corresponding years of birth and place (region or hospital) as the DS groups
† + higher than average, - lower than average
‡ Not statistically significant
§ 40 aborted fetuses with DS excluded
‖ Corrected for maternal age
¶ No information
** Corrected for post- and prematurity, delay caused by registration.

seasonal pattern at birth. Although a seasonal pattern in preterm birth may exist,[43] it would only cause a minor shift in a seasonal pattern for DS. Thus, it is unlikely that these factors would have masked a seasonal pattern in DS at birth. Moreover, some misclassification of Down syndrome might have occurred. Only 8 studies[24-28,31,38,39] mentioned the number of karyotyped cases of Down syndrome, the percentage of karyotyped cases varied from 15% to 100%. However, we do not expect that misclassification of Down syndrome was seasonally-bound and thus it can not have masked any association between season and the prevalence of Down syndrome at birth.

Another source of bias could be a seasonal pattern in spontaneous abortions. In their review, Hassold & Jacobs[44] reported that approximately 2.3 per cent of all spontaneous abortions, 1.3 per cent of all stillbirths and 0.13 per cent of all live births have trisomy 21. They estimated that trisomy 21 occurs in almost 0.45 per cent of all recognized pregnancies and that only 23.8 per cent of all conceptuses with trisomy 21 survive to term. As far as we know, there is only one article in which seasonality in DS was studied before birth. In that study, no seasonal variation was found in DS among 5,292 samples for prenatal diagnosis.[45] However, as there were only 45 DS cases, the results were not very reliable.

In brief, we have to conclude that there is no seasonal pattern in DS at birth. Thus we cannot support the hypothesis that the season influences the hormone production which results in Down syndrome. We cannot exclude the possibility, however, that a seasonal pattern in DS may exist at the time of conception, but disappears because of selective spontaneous abortions. To answer this question, a study using a very large number of prenatal karyotypes from an aselect group of women may provide more insight, especially if the nondisjunctions during the first meiosis of maternal origin are considered separately.

References

1. Bell JA. The epidemiology of Down's syndrome. Med J Aust 1991;155:115-117.
2. Polani PE, Briggs JH, Ford CE, Clarke CM, Berg JM. A mongol girl with 46 chromosomes. Lancet 1960;i:721-724.
3. Gaulden ME. Maternal age effect: The enigma of Down syndrome and other trisomic conditions. Mutation Research 1992;296:69-88.
4. Jongbloet PH. The effect of preovulatory overripeness of human eggs on development. In: Blandau RJ, (Ed.). Aging Gametes. Basel: S. Karger AG, The effect of preovulatory overripeness of human eggs on development. 1975;300-329.
5. Tamarkin L, Baird CJ, Almeida OFX. Melatonin: A coordinating signal for mammalian reproduction? Science 1985;227:714-720.
6. Timonen S, Franzas B, Wichmann K. Photosensibility of the human pituitary. Ann Chir Gynaec Fenn 1964;53:165-172.

7. Rameshkumar K, Thomas JA, Mohammed A. Atmospheric temperature & anovulation in south Indian women with primary infertility. Indian J Med Res 1992;96:27-28.
8. Levine RJ. Male factors contributing to the seasonality of human reproduction. Ann N Y Acad Sci 1994;709:29-45.
9. Weinberg CR, Moledor E, Baird DD, Wilcox AJ. Is there a seasonal pattern in risk of early pregnancy loss? Epidemiology 1994;5:484-489.
10. McDonald AD. Seasonal distribution of abortions. Br J Prev Soc Med 1971;25:222-224.
11. Sandahl B. A study of seasonal and secular trends in incidence of stillbirths and spontaneous abortions in Sweden. Acta Obstet Gynecol Scand 1974;53:251-257.
12. Kallan JE, Enneking EA. Seasonal patterns of spontaneous abortion. J Biosoc Sci 1992;24:71-75.
13. Lam DA, Miron JA. The seasonality of births in human populations. Ann Arbor, Michigan: Population Studies Center. University of Michigan; 1987.
14. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: I. Biology, sociology or both? J Biol Rhythm 1990;5:195-216.
15. Roenneberg T, Aschoff J. Annual rhythm of human reproduction: II. Environmental correlations. J Biol Rhythm 1990;5:217-239.
16. Leroy JG, Niermeijer MF. Klinische Genetica. In: Van den Brande JL, Van Gelderen HH, Monnens LAH, (Eds.). Kindergeneeskunde. Utrecht: Wetenschappelijke Uitgeverij Bunge, Klinische Genetica. 1990;141-158.
17. Antonarakis SE, Down Syndrome Collaborative Group. Parental origin of the extra chromosome in trisomy 21 as indicated by analysis of DNA polymorphisms. N Engl J Med 1991;324:872-876.
18. Sherman SL, Takaesu N, Freeman SB, *et al.* Trisomy 21: association between reduced recombination and nondisjunctions. Am J Hum Genet 1991;49:608-620.
19. Antonarakis SE, Petersen MB, McInnis MG, *et al.* The meiotic stage of nondisjunction in trisomy 21: determination by using DNA polymorphisms. Am J Hum Genet 1992;50:544-550.
20. Wehrung DA, Hay S. A study of seasonal incidence of congenital malformations in the United States. Br J Prev Soc Med 1970;24:24-32.
21. McDonald AD. Yearly and seasonal incidence of mongolism in Quebec. Teratology 1972;6:1-3.
22. Nielsen J, Petersen GB, Therkelsen AJ. Seasonal variation in the birth of children with aneuploid chromosome abnormalities. Humangenetik 1973;19:67-74.
23. Kaplan SD, Ament RP. Seasonal trend in Down's syndrome. PAS reporter 1975;13:1-5.
24. Lindsten J, Marsk L, Berglund K, *et al.* Incidence of Down's syndrome in Sweden during the years 1968-1977. Human Genet 1981;2 (suppl):195-210.
25. Gummere GR, Huether CA, Gartside PS. An analysis for temporal variation in Down syndrome births in Ohio, 1970-1979. Am J Hum Genet 1982;34:1003-1012.
26. Leisti J, Vahtola L, Linna SL, Herva R, Koskela SL, Vitali M. The incidence of Down syndrome in northern Finland with special reference to maternal age. Clin Genet 1985;27:252-257.
27. Baird PA, Sadovnick AD. Maternal age-specific rates for Down syndrome: changes over time. Am J Med Genet 1988;29:917-927.
28. Czeizel E. Some epidemiological characteristics of Down's syndrome in Hungary. Acta Morphol Hung 1988;36:63-77.
29. Källén B, Måsbäck A. Down syndrome. Seasonality and parity effects. Hereditas 1988;109:21-27.

30. Castilla EE, Orioli IM, Lugarinho R, *et al.* Monthly and seasonal variations in the frequency of congenital anomalies. Int J Epidemiol 1990;19:399-404.
31. Stoll C, Alembik Y, Dott B, Roth M-P. Epidemiology of Down syndrome in 118,265 consecutive births. Am J Med Genet 1990;7 (suppl):79-83.
32. Knox EG, Lancashire RJ. Epidemiology of congenital malformations. London: HMSO; 1991.
33. Collmann RD, Stoller A. A survey of mongoloid births in Victoria, Australia, 1942-1957. Am J Publ Hlth 1962;52:813-829.
34. Harlap S. A time-series analysis of the incidence of Down's syndrome in West Jerusalem. Am J Epidemiol 1974;99:210-217.
35. Rothman KJ, Fabia JJ. Place and time aspects of the occurrence of Down's syndrome. Am J Epidemiol 1976;103:560-564.
36. Holloway S, Emery AEH. Factors affecting the incidence of Down syndrome in Scotland. J Biosoc Sci 1977;9:453-465.
37. Leck I. Seasonality in Down syndrome. Lancet 1977;1:1057-1058.
38. Videbech P, Nielsen J. Chromosome abnormalities and season of birth. Human Genet 1984;65:221-231.
39. Kanai H, Nakamura I. Congenital malformations by month of birth. In: Miura T, (Ed.). Seasonality of birth. Progress in Biometeorology. Volume 6. Den Haag: SPB Academic Publishing, 12, Congenital malformations by month of birth. 1987;123-130.
40. Stark CR, Mantel N. Lack of seasonal- or temporal-spatial clustering of Down's syndrome births in Michigan. Am J Epidemiol 1967;86:199-213.
41. Boué J, Deluchat C, Nicolas H, Boué A. Prenatal losses of Trisomy 21. Human Genet 1981;2 (Suppl):183-193.
42. Paz JE, Otaño L, Gadow EC, Castilla EE. Previous miscarriage and stillbirth as risk factors for other unfavourable outcomes in the next pregnancy. Br J Obstet Gynaecol 1992;99:808-812.
43. Berkowitz GS, Papiernik E. Epidemiology of preterm birth. Epidemiol Rev 1993;15:414-443.
44. Hassold TJ, Jacobs PA. Trisomy in man. Annu Rev Genet 1984;18:69-97.
45. Drugan A, Bottoms SF, Johnson MP, Evans MI. Seasonal variation in conception does not appear to influence the rate of prenatal diagnosis of nondisjunction. Fetal Ther 1989;4:195-199.

3. Fecundity during treatment with in vitro fertilization

3.1. A more realistic approach to the cumulative pregnancy rate after in vitro fertilization

3.2. Prognostic models for the probability of achieving an ongoing pregnancy after in vitro fertilization and the importance of testing their predictive value

3.3. The search for externally valid prognostic models for ongoing pregnancy after in vitro fertilization

3.4. The impact of the woman's age on the success of standard and donor in vitro fertilization

3.1. A more realistic approach to the cumulative pregnancy rate after in vitro fertilization

A.M. Stolwijk, C.J.C.M. Hamilton, J.M.G. Hollanders, L.A. Bastiaans, G.A. Zielhuis

As most studies overestimate the cumulative pregnancy rate, a method is proposed to estimate a more realistic cumulative pregnancy rate by taking into account the reason for an early cessation of treatment with in vitro fertilization (IVF). Three methods for calculating cumulative pregnancy rates were compared. The first method assumed that those who stopped treatment had no chance at all of pregnancy. The second method, the one used most often, assumed the same probability of pregnancy for those who stopped as for those who continued. The third method assumed that only those who stopped treatment, because of a medical indication, had no chance at all of pregnancy and that the others who stopped had the same probability of pregnancy as those who continued treatment. Data were used from 616 women treated at the University Hospital Nijmegen, Nijmegen, the Netherlands. The cumulative pregnancy rates after five initiated IVF cycles for the three calculation methods were in the ranges from 37-51% for the positive pregnancy test result, 33-55% for a clinical pregnancy and 30-56% for an ongoing pregnancy. As expected, the first method underestimated the cumulative pregnancy rate and the second overestimated it. The third method produced the most realistic cumulative pregnancy rates.

Human Reproduction 1996;11:660-663

Introduction

Information about the probability of pregnancy after successive treatments with in vitro fertilization (IVF) is important for candidate patients and for the physicians who are counselling them. Several authors have reported cumulative pregnancy rates and recognized the importance of the variation in patient populations and treatment, including age, type of infertility and ovulation stimulation regimen, on the cumulative pregnancy rates.[1-6] Far less attention has been paid to the impact on cumulative pregnancy rated of the reason for early cessation of IVF treatment.[7-9]

As it is neither ethical nor practical to force patients to continue IVF treatments until pregnancy or for a fixed number of treatments (e.g. at least five), the real cumulative pregnancy rate after five IVF cycles for a specific population cannot be calculated. Therefore, to estimate the cumulative pregnancy rate, assumptions are necessary about the probability of the occurrence of pregnancy for those who discontinue treatment without achieving pregnancy. Most studies assume implicitly that all the patients who stop treatment early have the same probability of pregnancy as those who continue.[1-6,10] Haan *et al*[11] noted the importance of selective early cessation but compared the cumulative rate based on the assumption that the pregnancy rate of the first IVF treatment held good for the following IVF treatments with the cumulative rate based on the assumption that the same chance of pregnancy could be applied to women who stopped treatment early compared with those who continued treatment. Both of these methods will overestimate the real probability of pregnancy after successive IVF treatments. The other extreme, i.e. assuming that the women who stop IVF treatment early will never become pregnant, will obviously underestimate the cumulative pregnancy rate. The examples below show how large the bias can be.

Recently Tan *et al*[3] presented a cumulative live birth rate of 68.6% after five IVF treatments in women who had previously achieved an IVF live birth. A life-table analysis was used because it takes into account the experience of the entire cohort by using all the treatment cycles. However, this method implicitly assumes that the women who stopped IVF treatment early (in this case before the fifth treatment) had the same probability of having a live birth as those who continued treatment. Using the number of women per IVF treatment and the cumulative live birth rates, the calculated number of live births following each of the five IVF treatments were 21 live births in 105 women after the first treatment, six in 48 after the second, one in 30 after the third, one in 14 after the fourth and four in eight after the fifth. If we assume that none of the women who stop treatment early will ever achieve a live birth, the cumulative live birth rate would be 31%. The actual rate would be higher of course, but certainly not

as high as the 68.6% presented by Tan *et al.*

In the same manner, Guzick *et al*[1] calculated the cumulative rates for clinical pregnancy in treatments where oocyte retrieval was performed, excluding couples in whom the male partner had poor semen characteristics. They reported a cumulative pregnancy rate after six cycles of 59.6%. Furthermore, they predicted the cumulative pregnancy rates after nine and 12 cycles to be 75% and 84%, respectively. As all these calculations implicitly assumed the probability of pregnancy for those who stopped treatment early to be the same as for those who continued treatment, these rates are overestimated. Here, the assumption that those who stopped treatment early would not become pregnant leads to a cumulative pregnancy rate of 27% after six cycles. The actual rate would be between 27% and 59.6%, analogues to the first example.

More precise estimates of the cumulative pregnancy rates can be made if the reason for the early cessation of treatment is known. If, for instance, women stop treatment for financial or emotional reasons, they can be expected to have a higher probability of achieving a pregnancy than those who stop treatment because of poor IVF results such as low fertilization rates and poor embryo quality. To illustrate the importance of the assumption underlying the estimation of the cumulative pregnancy rates, here cumulative pregnancy rates have been calculated for patients of the University Hospital Nijmegen, the Netherlands during 1988-1993. Estimations assuming a difference in the probability of pregnancy for specific reasons for the discontinuation of treatment, were compared to those based on more extreme assumptions.

Materials and methods

At the University Hospital Nijmegen, the Netherlands, 872 women were treated for the first time with IVF between the August 1, 1988 and January 1, 1993. In this study only women were included who were treated with human menopausal gonadotropin in combination with a long protocol of gonadotrophin-releasing hormone agonist, with or without oral contraceptives during the preceding menstrual cycle, and who did not use donor spermatozoa (N = 616). Only the results of the first five initiated IVF treatments were analysed (whether or not oocyte aspiration and embryo transfer were performed). Three definitions of pregnancy were used: (i) positive pregnancy test result, measured 16 days after embryo transfer; (ii) clinical pregnancy — a positive pregnancy test result and ultrasonographic evidence of at least one gestational sac 4 weeks after embryo transfer; and (iii) ongoing pregnancy — a pregnancy continuing for >12 weeks after embryo transfer. To calculate the cumulative rates for each of these types of pregnancy, only the data up to the first pregnancy in question were included in the

analysis. To be clear, data were also included if women did not achieve pregnancy during the first five treatments.

Three assumptions were used to deal with the effect of the early cessation of IVF treatment (i.e. before a woman became pregnant): assumption I, women who stopped treatment had no chance of becoming pregnant; assumption II, women who stopped treatment had the same probability of becoming pregnant as those who continued; and assumption III, only the women who stopped treatment because of a medical indication had no chance of becoming pregnant, while the women who stopped treatment for other reasons had the same probability of pregnancy as those who continued treatment. A medical indication for stopping further IVF treatment was assumed to include: (i) a previous treatment with a fertilization rate of <10%, despite the presence of more than three large follicles (≥15 mm) on the day of human chorionic gonadotrophin (HCG) administration and the performance of oocyte aspiration, or (ii) three or less large follicles during two previous treatments.

The cumulative pregnancy rate after x initiated treatments was calculated as follows:[12] $[1 - \Pi (1 - \text{number of pregnant women in treatment } x / \text{number of women at risk in treatment } x)] \cdot 100\%$. In this formula, Π indicates the product of the terms specified within the parentheses for each of the x treatments.

Results

The number of women who achieved pregnancy per IVF treatment and who stopped after each treatment without becoming pregnant are shown in Table 1. Note, that the number of women who stopped treatment after a specific IVF cycle was the highest for the calculation of ongoing pregnancy, then for clinical pregnancy, and the lowest for the positive pregnancy test result. This was because all women who stopped treatment before achieving an ongoing pregnancy were also included in the numbers of those not achieving a positive pregnancy test result or a clinical pregnancy. The reverse situation did not apply. After the first IVF treatment, 42 women were advised to stop further treatment because of a medical indication; the corresponding figures were 58, 39 and 15 after the second, third and fourth treatments respectively.

The results of the calculations of the cumulative rates for a positive pregnancy test result, clinical pregnancy and ongoing pregnancy based on each of the three assumptions are shown in Figures 1A-C and Table 1. As expected, the cumulative pregnancy rates were the lowest with assumption I, highest with assumption II and intermediate with assumption III (Table 1). The cumulative pregnancy rates after five successive IVF treatments ranged from 37 to 51% for a positive pregnancy test result,

Table 1. Cumulative pregnancy rates for successive in vitro fertilization (IVF) treatments calculated on the basis of three assumptions

IVF treatment	No. of women who became pregnant	No. of women who stopped treatment*	No. of women at risk†			Cumulative pregnancy rate (%)		
			AI‡¶	AII§**	AIII‖††	AI	AII	AIII
Positive pregnancy test result								
1	131	120	616	616	616	21.3	21.3	21.3
2	56	102	485	365	407	30.4	33.3	32.1
3	34	133	429	207	307	35.9	44.3	39.6
4	5	31	395	40	179	36.7	51.3	41.3
5	0		390	4	158	36.7	51.3	41.3
Clinical pregnancy								
1	112	123	616	616	616	18.2	18.2	18.2
2	56	105	504	381	423	27.3	30.2	29.0
3	33	140	448	220	320	32.6	40.7	36.3
4	4	37	415	47	186	33.3	45.7	37.7
5	1		411	6	160	33.4	54.8	38.1
Ongoing pregnancy								
1	94	127	616	616	616	15.3	15.3	15.3
2	49	109	522	395	437	23.2	25.8	24.8
3	33	148	473	237	337	28.6	36.1	32.1
4	3	42	440	56	195	29.1	39.5	33.2
5	3		437	11	165	29.5	56.0	34.4

* Women who stopped after this treatment without becoming pregnant.

† Number of women who did not become pregnant in the previous IVF treatment(s).

‡ Assumption I (AI): women who stopped treatment had no chance of becoming pregnant.

§ Assumption II (AII): women who stopped treatment had the same probability of pregnancy as those who continued.

‖ Assumption III (AIII): women who stopped treatment because of a medical indication had no chance of becoming pregnant, while those who stopped treatment for other reasons had the same probability of pregnancy as those who continued.

¶ Women included who stopped treatment after the foregoing IVF treatment(s) without becoming pregnant.

** Women excluded who stopped treatment after the foregoing IVF treatment(s) without becoming pregnant.

†† Women who stopped treatment because of a medical indication included; women who stopped treatment for other reasons excluded (in comparison with those at risk in the case of assumption II, for IVF treatment 2, +42; for treatment 3, +100; for treatment 4, +139; and for treatment 5, +154).

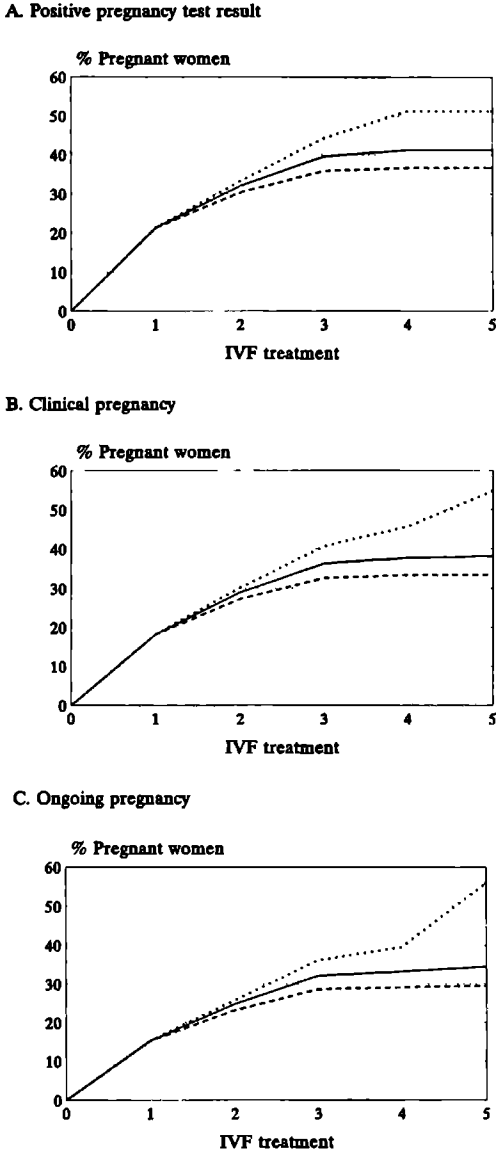


Figure 1. Cumulative rates for a positive pregnancy test result (A), clinical pregnancy (B) and ongoing pregnancy (C) for successive in vitro fertilization (IVF) treatments calculated on the basis of three assumptions: (- - -) assumption I, women who stopped treatment had no chance of becoming pregnant; (·····) assumption II, women who stopped treatment had the same probability of pregnancy as those who continued; (——) assumption III, women who stopped treatment because of a medical indication had no chance of becoming pregnant, while those who stopped for other reasons had the same probability of pregnancy as those who continued treatment.

from 33 to 55% for a clinical pregnancy and from 30 to 56% for an ongoing pregnancy. The cumulative pregnancy rates calculated by taking into account the reason for early cessation were 41% for a positive pregnancy test result, 38% for a clinical pregnancy and 34% for an ongoing pregnancy.

Discussion

In most of the studies that present cumulative pregnancy rates, the implicit assumption is that the women who stop IVF treatment before the occurrence of pregnancy have the same probability of becoming pregnant as those who continue. The consequence is an overestimation of the cumulative pregnancy rates, especially at higher IVF treatment numbers where many women stop further treatment. More realistic cumulative pregnancy rates can be estimated by incorporating the reason for early cessation of the IVF treatment in the assumption. Cumulative pregnancy rates based on the third assumption, which takes into account a medical indication for cessation of IVF treatment, will be the most reliable. However, as the probability of pregnancy in those with a medical indication for stopping treatment is assumed to be zero, it is reasonable to expect the cumulative pregnancy rate to be slightly underestimated. In particular, at higher treatment numbers this assumption will result in a slight underestimation, as women who received a medical indication for stopping further treatment only after the third or fourth cycle had reasonable results in at least one treatment cycle.

The deviations in the cumulative pregnancy rates found when using the different assumptions are highly dependent on the number of women who discontinued treatment before the occurrence of pregnancy. The overestimation of the real cumulative pregnancy rate by a life-table analysis (which uses the assumption that those who stop treatment have the same probability of becoming pregnant as those who continue) will be particularly large in clinics with a high percentage of patients for whom the IVF procedure is mainly performed for diagnostic purposes before a final decision is made to continue IVF treatment, e.g. in case of severe male infertility, unexplained infertility problems or a high level of follicle stimulating hormone in young women. In those cases, the method for the calculation of the cumulative pregnancy rate by taking into account whether or not one stopped treatment because of a medical indication (assumption III) is highly recommended.

In a recent study, Alsalili *et al*[13] calculated cumulative pregnancy rates by using a life-table analysis. They discussed the importance of the underlying assumption that the probability of pregnancy is the same for those who continue and those who discontinue treatment. They state that: 'The two factors, prognostic information ... ' (which is used

to decide whether or not to continue treatment) '... and successive cycle reduction in fertility, work in opposite directions on the assumption that pregnancy rates are constant for treated and non-treated individuals. Acknowledging its limitations, life-table analysis remains the conventional method for assessing IVF success and comparing results from different IVF centres.' Their argument for the two factors working in opposite directions is, however, incorrect. This can be explained by considering that those who stop treatment because of poor prior results (i.e. prognostic information) would have had a lower probability of pregnancy than those women who continued treatment, even if it is assumed that those who continued treatment would have a lower chance of pregnancy than in their previous cycle (i.e. successive cycle reduction). It seems that the authors confused the assumption made in life-table analysis: the probability of pregnancy for those who continue and those who discontinue treatment is not assumed to be the same at every cycle but in fact it applies to a specific cycle. Thus, the overestimation when using life-table analysis remains.

One result that might seem strange is that the estimated cumulative pregnancy rate after five treatments was higher for ongoing pregnancy than for clinical pregnancy and lowest for a positive pregnancy test result when using assumption II (i.e. women who stopped treatment had the same probability of pregnancy as those who continued), as shown in Table 1. However, this can be explained by looking at the denominator, i.e. the number of women at risk. All women who had not yet achieve a pregnancy but had stopped treatment are excluded from the denominator. As for the calculation of cumulative rates per type of pregnancy, women should only be included if they did not achieve that particular kind of pregnancy; notice that the numbers of women included at the fifth cycle are not the same for the three definitions of pregnancy. From Table 1 it can be inferred that only 11 women received a fifth IVF treatment because they had not achieved an ongoing pregnancy in the previous four cycles. Only six of them had not achieved a clinical pregnancy and four did not even have a positive pregnancy test result. As three and two of the women who were excluded from the calculation of the cumulative rate for a positive pregnancy test and clinical pregnancy after the fifth cycle, respectively, achieved an ongoing pregnancy during the fifth cycle, the cumulative pregnancy rate after the fifth cycle was higher for an ongoing pregnancy than for a clinical pregnancy or a positive pregnancy test result.

Comparison of the cumulative pregnancy rates between clinics and between types of assisted reproductive techniques can be very misleading. Reasons for this are not only the use of different definitions of pregnancy and the kind of assumption used to calculate the cumulative pregnancy rates, but also differences in the characteristics of

the populations (e.g. age, type of infertility, reproductive history), the characteristics of treatments (e.g. type of ovulation stimulation protocol, experience of the IVF clinic) and the number of and reason for couples discontinuing treatment before the occurrence of pregnancy. Caution is thus required when calculating, interpreting and comparing cumulative pregnancy rates. The figures presented here are only valid for the IVF clinic in Nijmegen, the Netherlands, between 1988 and 1993.

Multivariable prognostic models are necessary to take into account the influence of patient and treatment characteristics. In these models, the reason for cessation, i.e. a medical indication or another reason, should also be considered when calculating the cumulative pregnancy rate for specific patients. Only in that case can the cumulative pregnancy rate give reliable information for the candidate IVF patients.

References

1. Guzick DS, Wilkes C, Jones HW. Cumulative pregnancy rates for in vitro fertilization. Fertil Steril 1986;46:663-667.

2. Tan SL, Royston P, Campbell S, *et al.* Cumulative conception and livebirth rates after in-vitro fertilisation. Lancet 1992;339:1390-1394.

3. Tan SL, Doyle P, Maconochie N, *et al.* Pregnancy and birth rates of live infants after in vitro fertilization in women with and without previous in vitro fertilization pregnancies: a study of eight thousand cycles at one center. Am J Obstet Gynecol 1994;170:34-40.

4. Tan SL, Maconochie N, Doyle P, *et al.* Cumulative conception and live-birth rates after in vitro fertilization with and without the use of long, short, and ultrashort regimens of the gonadotropin-releasing hormone agonist buserelin. Am J Obstet Gynecol 1994;171:513-520.

5. Simon A, Ronit C, Lewin A, Mordel N, Zajicek G, Laufer N. Conception rate after in vitro fertilization in patients who conceived in a previous cycle. Fertil Steril 1993;59:343-347.

6. Check JH, Baker A, Lurie D, Benfer K, Callan C. Comparison of the cumulative probability of pregnancy after in vitro fertilization-embryo transfer by infertility factor and age. Fertil Steril 1994;61:257-261.

7. Page H. Calculating the effectiveness of in-vitro fertilization. A review. Br J Obstet Gynaecol 1989;96:334-339.

8. Te Velde ER, Koudstaal J, Eimers JM. Assisted conception for infertility (letter). Br Med J 1992;305:1097-1098.

9. De Mouzon J, Bachelot A, Spira A. Establishing a national in vitro fertilization registry: methodological problems and analysis of success rates. Stat Med 1993;12:39-50.

10. Hull MGR, Eddowes HA, Fahy U, *et al.* Expectations of assisted conception for infertility. Br Med J 1992;304:1465-1469.

11. Haan G, Bernardus RE, Hollanders HMG, Leerentveld BA, Prak FM, Naaktgeboren N. Selective drop-out in successive in-vitro fertilization attempts: the pendulum danger. Hum Reprod 1991;6:939-943.

12. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York, NY, USA: John Wiley & Sons; 1980.

13. Alsalili M, Yuzpe A, Tummon I, *et al.* Cumulative pregnancy rates and pregnancy outcome after in-vitro fertilization: >5000 cycles at one center. Hum Reprod 1995;10:470-474.

3.2. Prognostic models for the probability of achieving an ongoing pregnancy after in vitro fertilization and the importance of testing their predictive value

A.M. Stolwijk, G.A. Zielhuis, C.J.C.M. Hamilton, H. Straatman, J.M.G. Hollanders, H.J.M. Goverde, P.A. van Dop, A.L.M. Verbeek

The aim of this study was to create reliable models to predict the probability of achieving an ongoing pregnancy during in vitro fertilization (IVF) treatment: model A, at the start of the first treatment, model B, at the time of embryo transfer, and model C, during the second treatment at the end of the first IVF treatment. Prognostic models were created using data from the University Hospital Nijmegen (N = 757) and applied to the data from the Catharina Hospital Eindhoven (N = 432), the Netherlands, to test their predictive performance. The predictions of model B (made at time of embryo transfer) were fairly good (c = 0.672 in the test population). For instance, 93% of the patients who had a predicted probability of achieving an ongoing pregnancy of <10% did not achieve an ongoing pregnancy. However, the predictions of the other two models (A and C) for Eindhoven were less reliable. The predictive value of model C was fairly high in Nijmegen (c = 0.673). Its poor performance in the test population may be explained partly by differences in effectiveness of the ovulation stimulation protocols and the decision about when to discontinue the cycle. Thus, before using prognostic models at an IVF centre, their reliability at that specific centre should be tested.

Human Reproduction 1996;11:2298-2303

Introduction

The probability that a patient will achieve an ongoing pregnancy should be evaluated as accurately as possible before a patient enters a programme for in vitro fertilization (IVF) and during the course of her treatment with IVF. In addition to the age of the woman and the aetiology of infertility (the standard indicators for success), better rules would be welcome for physicians when counselling a patient. Potential predictors of IVF success are: patient characteristics at entry to the programme, characteristics of the treatment itself and during treatment, and the intermediate results.

Most studies on factors that may predict pregnancy after treatment with IVF have investigated only a few indicators, for instance age and the type of infertility,[1-3] baseline follicle stimulating hormone (FSH), luteinizing hormone (LH) and oestradiol concentrations,[4] ovulation stimulation treatment and ovarian response,[5] endometrium thickness and uterine artery flow,[6] sperm characteristics[7] and age, oestradiol concentration, number and quality of oocytes and embryos.[8] However, various simultaneous factors may influence the probability of achieving an ongoing pregnancy after IVF. It would therefore be desirable to create a model to predict the probability of achieving an ongoing pregnancy which includes all the relevant factors. Until now, only a few attempts have been made to do this for IVF[9,10] and other assisted reproductive techniques.[11,12] In the study by Hughes *et al*,[9] age and failed fertilization due to poor sperm quality had a predictive value for success in subsequent IVF cycles. Haan *et al*[10] found that the probability of achieving an ongoing pregnancy after IVF treatment was increased by the presence of idiopathic infertility and decreased by the presence of a male factor, one ovary, the woman's age ≥36 years, primary infertility of at least 5 years duration and by a higher number of previous IVF treatments. Multivariate prognostic models should not be confused with explanatory models such as recently published by Roseboom *et al*.[13] They discussed a multivariate model to explain the variation in the probability of pregnancy after embryo transfer. The variation was explained by the woman's age, average embryo morphology score, number of transferred embryos and an interaction term between tubal pathology and the woman's age. However, exclusion of the main effect of tubal pathology in the model makes a meaningful interpretation of the multivariate model difficult[14] and may cause bias.[15] Moreover, their statement in the results section '... with a 1 year increase of age, the probability of pregnancy for non-tubal patients decreased by 21%...' is obviously mistaken as a result of a wrong interpretation of the odds ratio in their study. Critical remarks can also be made about the methods used in the other four studies mentioned above.[9-12] All the cycles were combined, irrespective of the number of previous IVF treatments and the number of treatments per patient. Some studies based the inclusion

of factors on statistical significance of the relationship in univariate analyses, which can be influenced by other factors, instead of on the increase in the predictive power in multivariate models. Moreover, the predictions of these models were never tested in other populations. Thus, the validity of these prognostic models when used at other IVF centres can be questioned.

The purpose of this study is to create reliable models to predict the probability of achieving an ongoing pregnancy during the first or second treatment cycles with IVF. We used data from the University Hospital Nijmegen, the Netherlands, to develop the models, and data from another centre to test their predictive value.

Materials and methods

To develop the prognostic models, data were used from couples who were treated by IVF for the first time in the period March 1991 to January 1995 at the University Hospital Nijmegen, the Netherlands. During this period IVF treatment hardly changed. To test these models, data were used from the Catharina Hospital, Eindhoven, the Netherlands. Guidelines on indications for IVF treatment in the Netherlands have been described by Jansen.[16] In short, couples are only offered IVF treatment in case of bilateral tubapathology, in cases of unilateral tubapathology, male factor, endometriosis or cervical factor when other infertility treatments had not resolved the problem, and in case of idiopathic infertility after an infertility duration of at least 3 years. For both populations data were only included if the complete IVF treatment had been carried out at that particular IVF centre, no donor oocytes had been used and no intra-cytoplasmatic sperm injection (ICSI) had been performed. Patient characteristics prior to treatment are given in Table 1.

Ongoing pregnancy was defined as a pregnancy which continued for longer than 12 weeks after embryo transfer. To predict the probability of achieving an ongoing pregnancy, three models were developed that employed different moments of prediction. To predict the probability of achieving an ongoing pregnancy during the first IVF treatment, model A was made at the start and model B at the time of embryo transfer. For the prognosis of achieving an ongoing pregnancy during the second IVF treatment, model C was created at the end of the first IVF treatment. Table 2 presents the number of patients and pregnancies at each prediction moment.

Model A

This model was based on predictions made at the start of the first IVF cycle regarding the probability of achieving an ongoing pregnancy during the first IVF cycle. To develop this model, data were available from 757 couples whose first IVF cycle took place in

Table 1. Patient characteristics of the populations at the start of the first in vitro fertilization (IVF) cycle

	Nijmegen (N = 757)					Eindhoven (N = 432)				
	Min.	Max.	Mean	SD	Median	Min.	Max.	Mean	SD	Median
Woman's age (years)	22	47	32.9	4.0	33	21	43	31.8	4.1	32
Duration of infertility (years)*	0	20.5	4.4	2.8	4.0	0	20.5	3.7	2.7	3.5
Basal FSH (IU/l)	<0.6	23	6.1	2.8	5.7	§				
		n		%†			n		%‡	
≥1 Preceding gestations		256		33.8			162		37.5	
≥1 Preceding spontaneous abortions		126		16.6			29		8.3	
≥1 Preceding ectopic pregnancies		62		8.2			23		6.6	
≥1 Preceding deliveries		138		18.2			50		14.2	
Indication for IVF										
Tubal exclusively		168		22.2			147		34.6	
Tubal and other(s)		137		18.1			42		9.9	
Male factor exclusively		133		17.6			94‖		22.1	
Male factor and other(s)		190		25.1			34¶		8.0	
Endometriosis exclusively		44		5.8			34		8.0	
Endometriosis and other(s)		119		15.7			33		7.8	
Cervical factor exclusively		27		3.6			1		0.2	
Cervical factor and other(s)		91		12.0			1		0.2	
Idiopathic infertility		138		18.2			96		22.6	
Two ovaries		708		93.7			395		93.4	
Sperm characteristics										
≥20 * 10^6/ml		630		83.2			§			
≥60% Normal forms		433		57.2			§			
≥50% Motile		416		55.0			§			
Quality of motility ≥4**		660		87.2			§			
Anti-sperm antibodies, ♂ or ♀		66		8.7			10		2.3	
In sperm		38		5.0			§			
In woman's serum		29		3.8			§			
Use of donor spermatozoa		0		0.0			4		0.9	

* Number of missing values for duration of infertility in Nijmegen n = 374.

† Number of missing values in Nijmegen: for two ovaries n = 1, anti-sperm antibodies ♂ or ♀, and in sperm n = 2.

‡ Number of missing values in Eindhoven: for ≥1 preceding spontaneous abortions, ectopic pregnancies, deliveries, respectively n = 81, 81, 80, for the indications of IVF n = 7, for two ovaries n = 9.

§ No information available.

‖ Donor spermatozoa was used for three patients.

¶ Donor spermatozoa was used for one patient (the other indication for IVF was tubal factor).

** On a scale from 1 (worst) to 5 (best).

Nijmegen. To induce ovulation, all the patients received a long protocol of gonadotrophin-releasing hormone (GnRH) agonist (usually Leuprolide; Abbott B.V., Amstelveen, the Netherlands or Suprefact; Hoechst Holland N.V., Amsterdam, the Netherlands) that was started on day 21 of the previous cycle and human menopausal gonadotrophin (HMG, Humegon; Organon Int. B.V., Oss, the Netherlands). Additionally, from August 1991 to January 1994, all the patients received oral contraceptives during the cycle that preceded the IVF cycle. To improve synchronization of follicle growth, some women received oral contraceptives before or after this period. To test the model, data were available from 432 couples from Eindhoven who underwent their first IVF treatment between January 1990 and June 1995 (another five couples were excluded from this population because information about the occurrence of an ongoing pregnancy was lacking). In this test population, the type of ovulation induction used most often (92.2%) was a short protocol of GnRH agonist (usually Suprefact; Hoechst Holland N.V.) and HMG (Humegon; Organon Int. B.V.) in a few cases supplemented by progestins in the preceding cycle.

Table 2. Number of patients and ongoing pregnancies

	Nijmegen			Eindhoven		
	N	Pregnancies		N	Pregnancies	
		n	%		n	%
At start of first IVF	757	88	11.6	432	46	10.6
At embryo transfer of first IVF	604	88	14.6	300	46	15.3
At start of second IVF	454	61	13.4	275	29	10.5

Potential prognostic factors for the model that employed the onset of the first IVF cycle as the moment of prediction could only consist of information known at that moment, i.e. patient characteristics: age, period of infertility, reproductive history, basal FSH, indication(s) for IVF treatment, one or both ovaries present, sperm characteristics, anti-sperm antibodies in the woman or man, and information about the treatment protocol being used at that time: type of hormonal ovulation stimulation, maximum number of embryos that would be transferred, timing of human chorionic gonadotrophin (HCG) administration and type of culture medium. In Nijmegen, data on the duration of infertility were only available from patients who started IVF treatment between 1993-1994. Therefore the effect of the duration of infertility could only be estimated using the data from these 383 couples. Donor spermatozoa had not been used in Nijmegen, but it had been used in the test population in four and six patients during the first and second IVF cycles, respectively. If donor spermatozoa had been used, the sperm characteristics were considered to be good and the indication for IVF 'male factor' was considered to be

absent. We disregarded the results of cryopreserved embryo transfer.

Model B

Based on predictions made at the time of embryo transfer regarding the probability of achieving an ongoing pregnancy during the first IVF cycle. Only the data from couples who underwent embryo transfer during the first cycle were used to develop this model. Data were available from 604 (79.8% of the 757) couples from Nijmegen. To test the model, data could be used from 300 (69.4% of the 432) couples from Eindhoven. At this moment, information was added about preceding events during the cycle as potential prognostic factors, i.e. quality and number of oocytes retrieved, number of oocytes fertilized, quality and number of embryos transferred and whether the transfer had been uncomplicated as indicated by the use of a Wallace catheter, because in difficult cases a stiffer, Frydman catheter was used. In addition, information was known about the experience of the physician who performed the puncture and transfer; this could be used as a potential prognostic factor. Again, the results of cryopreserved embryo transfer were disregarded.

Model C

Based on predictions made at the end of the first IVF cycle regarding the probability of achieving an ongoing pregnancy during the second IVF cycle. To create this model, data were used from couples who did not have an ongoing pregnancy after the first IVF cycle or after a transfer of cryopreserved embryos and who started a second IVF cycle. In Nijmegen and in Eindhoven, 454 and 278 couples started a second IVF cycle, respectively. In Eindhoven, information about ongoing pregnancy was lacking for three couples during the second cycle, so the data from 275 couples could be used for the test. In addition to the factors mentioned above, the pregnancy test result after the first IVF cycle was a potential prognostic factor in this model.

Statistical analysis

Models were developed by using logistic regression analysis. The first step was to develop a prognostic model based on patient characteristics and, if appropriate, the intermediate IVF treatment results. The second step was to evaluate whether treatment characteristics added any prognostic value to the model. The third step was to test the model.

Criteria for accepting variables as predictive factors in the model were based on statistical significance and added prognostic value, evaluated by using the c index [i.e. (number of concordant pairs + 0.5 * the number of tied pairs)/ total number of pairs] .[17,18] The c can be interpreted as the probability of a correct prediction for a random pair of a woman with an ongoing pregnancy and a women without a pregnancy. It is equal to the area under a receiver operating characteristic (ROC) curve.[19] For the development of a

prognostic model, the erroneous exclusion of any prognostic factors (because of too little power) would be more deleterious than including too many factors. Therefore these criteria were given a high and low cut-off point, respectively; P value < 0.10 and c > 0.005. The variables were selected according to a method akin to a stepwise selection method. Here, the selection criteria is based not only on a P value (<0.10), but also on a change in c (>0.005). Special attention was given to multicollinearity. If this was present, only the variable with the highest predictive power was included in the multivariate model. If a variable did not meet the criteria in a univariate analysis, it thus could still be included in the prognostic model if the variable met the criteria when it was included in a multivariate model, i.e. after taking into account the prognostic value of other variables. In addition, a variable was omitted from the model if another factor was a stronger predictor and showed no additional predictive value. For sperm characteristics combined variables were created and their predictive value was evaluated against that of the separate sperm characteristics.

To test the predictive validity of the models, the data from the other centre were applied. As the data from Nijmegen contained more potential predictors than the data from Eindhoven, the models selected as the best predictive could not always be fully tested. If a specific variable was lacking, the model was modified, if possible, by exchanging it with a similar variable, or otherwise by excluding the variable. To evaluate the reliability of the model, the c was calculated. If the model had reasonable prognostic value, the predicted probability and the observed result of IVF were compared.

Results

The models for predicting the probability of achieving an ongoing pregnancy, developed with the data from Nijmegen and tested with the data from Eindhoven, are presented in Table 3.

Model A

During the first IVF cycle, 88 (11.6%) out of the 757 women from Nijmegen and 46 (10.6%) out of the 432 women from Eindhoven achieved an ongoing pregnancy. The only factors that had predictive value were a previous gestation and the woman's age. During testing, this model did not show any predictive value when applied to the data from Eindhoven (c = 0.497).

Model B

Embryo transfer was performed in 604 (79.8%) out of the 757 couples from Nijmegen in the first IVF cycle. In Eindhoven, embryo transfer was performed in 300 (69.4%) out of the 432 couples. The ongoing pregnancy rate per transfer was 14.6% in Nijmegen and

Table 3. Prognostic models for the probability of achieving an ongoing pregnancy (P) during the first in vitro fertilization (IVF) cycle or second IVF cycle

Model	Ln [P/(1-P)] =	SE(ß)	-2ln(L_1/L_2) df p value	Nijmegen at development		Eindhoven at testing	
				N*	c	N*	c
A	− 0.3350	0.9503	14.04	757	0.612	431	0.497
	+ 0.8151 * ≥1 preceding gestation	0.2349	df=2				
	− 0.0620 * woman's age (years)	0.0297	p=0.0009				
B	− 4.2034	0.5399	48.96	603	0.721	171	0.672
	+ 0.5290 * ≥1 preceding gestation	0.2422	df=4				
	+ 0.0630 * no. fertilized oocytes	0.0260	p=0.0001				
	+ 0.3464 * no. transferred embryos	0.1711					
	+ 0.4377 * no. transferred embryos of at least good quality	0.1297					
C	− 4.0236	0.7812	20.88	454	0.673	271	0.528
	+ 0.9886 * woman's age ≤30 years	0.4146	df=4				
	+ 0.6001 * woman's age 31-35 years	0.3886	p=0.0003				
	− 0.8412 * idiopathic infertility	0.4537					
	+ 1.8638 * embryo transfer during first IVF cycle	0.7336					

* Patients with missing values on one or more of the variables were excluded, i.e. for Nijmegen 0, 1 and 0, and for Eindhoven 1, 129 and 4 for models A (prediction at start of first IVF cycle regarding probability during first cycle), model B (prediction at embryo transfer regarding probability during first cycle) and model C (prediction at end of first IVF cycle regarding probability during second cycle), respectively.

15.3% in Eindhoven. The prognostic model included the factors: at least one preceding gestation, the number of fertilized oocytes, the number of transferred embryos and the number of transferred embryos of at least good quality. The probability of achieving an ongoing pregnancy increased if there had been a preceding gestation and the higher the numbers. During the test, this model showed good predictive value (c = 0.672) and good predictive performance, as shown in Table 4. For instance, 93% of the women with a predicted probability of achieving an ongoing pregnancy of <10% did not achieve an ongoing pregnancy after embryo transfer.

Model C

To predict the probability of achieving an ongoing pregnancy during the second IVF cycle, only the data from the couples who received a second treatment could be used. Of the 454 couples who received a second IVF treatment in Nijmegen, 61 (13.4%) achieved an ongoing pregnancy. In Eindhoven this occurred in 29 (10.5%) out of the 275 couples who underwent a second IVF treatment. The best prognostic model is shown in Table 3. Of prognostic value were: the woman's age in age-groups, the presence of idiopathic

infertility and embryo transfer during the first IVF cycle. However, this model did not show any predictive value in the test population (c = 0.528).

Table 4. Predicted and observed percentages and numbers of women with an ongoing pregnancy during the first in vitro fertilization (IVF) treatment at the time of embryo transfer in Eindhoven

	Predicted probability (%)							
Observed	0-<5	5-<10	10-<15	15-<20	20-<25	25-<30	≥30	Total
Percentage ongoing pregnancy	0	8	16	19	23	25	50	15
No. women pregnant	0	5	6	6	3	2	3	25
Total number of women	16	60	37	31	13	8	6	171*

* No prediction could be made for 129 women, because no information was available about the number of transferred embryos of at least good quality.

Discussion

This study showed that models for prediction of ongoing pregnancy due to IVF treatment can be developed with a fairly high prognostic value. However, this does not imply that the same models are predictive for patients treated at another clinic or even at the same clinic. Of the three models, only the one that made a prediction at the time of embryo transfer was fairly reliable in the other population. The other two models that made predictions at the start of treatment or after the first IVF cycle, however, seemed to be of little value when used in Eindhoven. Although model B, which predicts at time of embryo transfer, is of little clinical importance, it gives information about the reasons for the inadequacy of the prediction at the start of the cycle. For the two models at the start of the cycle, the ovarian response and oocyte aspiration are very important, but cannot be included as prognostic factors in the models because this information is not available at the start of the treatment. Whereas in the model that made a prediction at time of embryo transfer, the number and quality of the retrieved oocytes are potential prognostic factors. Therefore, one explanation for the poor reliability might be differences in the effectiveness of the ovulation simulation protocols, the long protocol of GnRH agonist in Nijmegen and the short protocol in Eindhoven. No oocyte aspiration was performed during the first IVF treatment in 7.4% (56 out of the 757) and 21.8% (94 out of the 432) of the women from Nijmegen and Eindhoven, respectively. During the second IVF treatment, these percentages were 4.6% (21 out of the 454) and 15.3% (42 out of the 275), respectively. Not only might the effectiveness of the ovulation stimulation protocol have influenced the cancellation rate, but also the timing of this decision differed between the two centres.

During the first IVF cycle, the percentage of cancelled cycles for the reason of too many follicles was only 1.8% in Nijmegen, but was as high as 31.5% in Eindhoven. This decision was made in Nijmegen if >25 follicles were present in combination with an oestradiol concentration of >20,000 pmol/l, whereas in Eindhoven, cycles were cancelled when >20 follicles were present. Whether the models developed in Nijmegen can make more accurate predictions if they are applied to an IVF centre that uses a long protocol of GnRH agonist and with fewer cancelled cycles remains to be seen.

The present models were adapted to make testing possible, given the information available in the test population. The changes were negligible. Models A and C were not changed at all. In model B the number of follicles >15 mm was initially included in the model, but because of lack of this information in the Eindhoven population, it was exchanged with the number of fertilized oocytes. Moreover, in model B the sperm characteristics <60% normal forms and/or $<20*10^6$ spermatozoa per ml added minor predicting value, and were excluded from the model. Note that basal FSH had no additional predictive value, nor had the indications for IVF, except for idiopathic infertility in model C.

For prognosis, the predictive value of a positive test and of a negative test are of more practical value than the sensitivity and specificity of a test. The predictive value of a positive test is the proportion of patients with a positive test who achieve an ongoing pregnancy, and the predictive value of a negative test is the proportion of the patients with a negative test who do not achieve an ongoing pregnancy. Thus, they illustrate whether the prognosis was right. Whereas the sensitivity and specificity of a test indicate whether the patients who achieved an ongoing pregnancy were classified well by the test. All these measures can be easily calculated using the data of Table 4. For instance, assume the cut-off point for the test to be a predicted probability of 5%; the test is positive if the predicted probability is ≥5% and negative if <5%. The positive predictive value of this test is 16% (25/155) and the negative predictive value is 100% (16/16). This demonstrates that the test can indicate patients who do not achieve an ongoing pregnancy after IVF, but cannot predict who achieves an ongoing pregnancy. The sensitivity and specificity of this test are 100% (25/25) and 11% (16/146), respectively.

Obviously, clinicians select their patients before treatment with IVF. If the study populations had included more extreme groups, those with a very high or a very low probability of success, then the reliability of the prognosis would have been better. The models we created only apply to populations that lie within the range of the characteristics presented in Table 1. As women of 40 years of age or older were poorly represented in Nijmegen (n = 34), the models may not be valid for them. In addition, information on the

duration of infertility was only available from 383 patients in Nijmegen. The potential prognostic effect of the duration of infertility might not have been detected because of too few observations.

As the data were gathered retrospectively, it was not always possible to obtain full sets of information from the two databases. In some cases data were missing, or they were not present in the desired form. Moreover, the two hospitals had their own method of performing IVF and the patient populations might have differed on other aspects than those studied. Therefore it was more difficult to create a model that would make reliable predictions than if the data had been gathered in a standardized way for the purpose of prognostic studies at hospitals which use the same treatment protocols and the same definitions for each variable. To make it possible to create reliable prognostic models, we recommend setting up uniform national registries which also contain information about the basic fertility workup.

The importance of testing prognostic models is evident. Untested prognostic models can be worthless when used for prediction at another (or possibly even the same) IVF clinic. Before a model can be used by another IVF centre, it should be tested with retrospective data from that centre, to establish whether it is a predictive model in that centre. Even before a model is implemented in the centre were it was developed, it should be tested with an entirely separate set of data from the same centre before one can rely on its predictive properties.

References

1. Piette C, De Mouzon J, Bachelot A, Spira A. In-vitro fertilization: influence of women's age on pregnancy rates. Hum Reprod 1990;5:56-59.

2. Hull MGR, Eddowes HA, Fahy U, *et al.* Expectations of assisted conception for infertility. Br Med J 1992;304:1465-1469.

3. Check JH, Lurie D, Dietterich C, Callan C, Baker A. Adverse effect of a homogeneous hyperechogenic endometrial sonographic pattern, despite adequate endometrial thickness on pregnancy rates following in-vitro fertilization. Hum Reprod 1993;8:1293-1296.

4. Padilla SL, Bayati J, Garcia JE. Prognostic value of the early serum estradiol response to leuprolide acetate in in vitro fertilization. Fertil Steril 1990;53:288-294.

5. Dor J, Seidman DS, Ben Shlomo I, Levran D, Karasik A, Mashiach S. The prognostic importance of the number of oocytes retrieved and estradiol levels in poor and normal responders in in vitro fertilization (IVF) treatment. J Assist Reprod Genet 1992;9:228-232.

6. Spernol R, Hecher K, Schwarzgruber J, Szalay S. Doppler-Flow-Messungen in der Ateria uterina. Ein Prognosefaktor für den Erfolg bei der Behandlung durch IVF? Ultraschall Med 1993;14:175-177.

7. Enginsu ME, Pieters MH, Dumoulin JC, Evers JL, Geraedts JP. Male factor as determinant of in-vitro fertilization outcome. Hum Reprod 1992;7:1136-1140.

8. Fluker MR, Siu CK, Gunby J, Daya S. Cycle characteristics and outcome in relation to ovarian response during in vitro fertilization. J Assist Reprod Genet 1993;10:504-512.

9. Hughes EG, King C, Wood EC. A prospective study of prognostic factors in in vitro fertilization and embryo transfer. Fertil Steril 1989;51:838-844.
10. Haan G, Bernardus RE, Hollanders JMG, Leerentveld RA, Prak FM, Naaktgeboren N. Results of IVF from a prospective multicentre study. Hum Reprod 1991;6:805-810.
11. Guzick DS, Balmaceda JP, Ord T, Asch RH. The importance of egg and sperm factors in predicting the likelihood of pregnancy from gamete intrafallopian transfer. Fertil Steril 1989;52:795-800.
12. Nelson JR, Huppert L, Corson SL, *et al.* Predicting success of gamete intrafallopian transfer. Fertil Steril 1993;60:116-122.
13. Roseboom TJ, Vermeiden JPW, Schoute E, Lens JW, Schats R. The probability of pregnancy after embryo transfer is affected by age of the patient, cause of infertility, number of embryos transferred and the average morphology score, as revealed by multiple logistic regression analysis. Hum Reprod 1995;10:3035-3041.
14. Breslow NE, Day NE. Statistical methods in cancer research. Volume 1 - The analysis of case-control studies. Lyon: International Agency for Research on Cancer; 1980.
15. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Principles and quantitative methods. New York: Van Nostrand Reinhold Company; 1982.
16. Jansen CAM. Indicaties voor in vitro fertilisatie. Nederlands Tijdschrift voor Obstetrie & Gynaecologie 1993;106:35-36.
17. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247:2543-2646.
18. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-387.
19. Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology 1982;143:29-36.

3.3. The search for externally valid prognostic models for ongoing pregnancy after in vitro fertilization

A.M. Stolwijk, H. Straatman, G.A. Zielhuis, C.A.M. Jansen, D.D.M. Braat, P.A. van Dop, A.L.M. Verbeek

This study aimed to make reliable prognostic models for predicting ongoing pregnancy after the first and second in vitro fertilization (IVF) cycles. Models were developed using data from the University Hospital Nijmegen 1991-1994 and tested using data from Nijmegen 1995-1996 and two other centres. The internal validity was estimated by bootstrapping. Predictions were calibrated by shrinkage and prior probability. Discrimination was studied by the c index and observed proportions of women with a low predicted probability. Models, whether or not after shrinkage, did not discriminate well. Calibration by prior probability did not show any advantage. Nevertheless, specific situations seemed to exist in which the model for ongoing pregnancy during the second IVF cycle was reliable. The predictions were inaccurate in the Nijmegen 1995-1996 population because of too small numbers of patients with a low probability. In general, the prognostic models should not yet be implemented in clinical practice. Further development and testing of the models is warranted.

(submitted)

Introduction

Prognostic models can facilitate decision-making for patients and physicians and are therefore becoming increasingly popular. Important prerequisites for the implementation of these models in clinical practice are that they not only have adequate internal validity, but also show high external validity. Testing the internal validity using, for instance, the split-half method, gives only an indication of the amount of overfitting of the model to the data at hand. Whether an internally valid model will predict well in another population, thus whether it is externally valid, is still the question. To examine this, a model should be applied to other data than those on which the model was based. The predictive accuracy of a prognostic model can be expressed by calibration and discrimination.[1] Calibration refers to the amount of bias in the predictions, while discrimination refers to the ability to separate patients with different outcomes. Unfortunately, prognostic models for the probability of pregnancy presented in the literature have not been validated at all[2-5] or only their internal validity has been checked.[6,7] The conclusions based on these studies are not always in line with the uncertainty still present in the models, however. For instance, Collins *et al*[7] estimated only the internal validity of their model, but concluded that the model predicted sufficiently accurately to be useful in the clinical management of infertility. Hughes *et al*[3] concluded that their prognostic model '... provides patients with a more accurate prognosis before treatment', although their model was not even internally validated. Nelson *et al*[5] based their conclusions regarding the correctness of the predictive performance of their model on a comparison between the observed outcomes and predicted outcomes in the same - total - population. Appropriately, Eimers *et al*[6] mentioned that their model should be applied to an external population to check its validity. Nevertheless, they concluded that '...the chance to conceive spontaneously can be predicted.'

Recently, we have developed two models to predict the probability of achieving an ongoing pregnancy during the first and second in vitro fertilization (IVF) cycles.[8] The actual moments of prediction were the start and the end of the first cycle, respectively. We used data from one centre for developing the models and data from another centre for testing them. The models did not predict well at the test centre. The aims of this study were to determine whether this was just bad luck or a real reflection of low validity, and to explore possible ways of increasing the validity of these prognostic models for ongoing pregnancy after IVF treatment. To achieve these goals, we tested the original models using data from two other populations. Corrections were made for obvious differences between protocols for cancelling treatment by means of restriction.

Moreover, we studied whether the two developed models were overfitted because of the inclusion of too many potential prognostic factors. Internal validity was further evaluated by bootstrapping. Shrinkage was used to calibrate the predictions in order to avoid the effects of overfitting of the model. Further calibration was performed by taking the prior probability into account.

Materials and Methods

Two prognostic models were developed: Model I for predicting the probability of achieving an ongoing pregnancy during the first IVF treatment, and Model II for predicting the probability of achieving an ongoing pregnancy during the second IVF treatment; the first prediction was made at the start of the first IVF treatment, the latter at the end of the first treatment. Ongoing pregnancy was defined as a pregnancy which continued for longer than 12 weeks after embryo transfer. To develop and test the prognostic models, data from patients were only included if the complete IVF treatment had been carried out at that particular IVF centre, hormonal ovulation induction had been performed, no donor oocytes had been used and no intra-cytoplasmatic sperm injection had been performed. In addition to these criteria, which were also employed in our previous study, we excluded any cycles in which donor sperm had been used and defined a stricter criterion for idiopathic infertility (excluding patients who underwent IVF treatment only because of hormonal dysfunction or unsuccessful intrauterine insemination by donor sperm). To develop Model II, for predicting an ongoing pregnancy during the second treatment, only data from patients were included who underwent a second IVF cycle after an unsuccessful first IVF treatment. The results of cryopreserved embryo transfer were disregarded. Four databases were available from Dutch IVF centres:

1. University Hospital Nijmegen, 1991-1994

Data from this population were used to develop the prognostic models described previously. The population consisted of 757 couples who were treated for the first time with IVF. Data from 454 of them could be used for predicting ongoing pregnancy during the second IVF cycle.

2. University Hospital Nijmegen, 1995-1996

After the development of the prognostic models, new data from the same IVF centre became available. These concerned data from 208 patients who underwent IVF for the first time and 127 patients who underwent their second IVF cycle during this period.

3. Catharina Hospital Eindhoven

Data from patients in Eindhoven were used to test the models developed in our

previous study. Data were available from 428 couples who underwent their first IVF treatment between 1990 and 1995 and from 268 couples who underwent a second treatment.

4. Diaconessenhuis Voorburg

Data were available from 1424 patients who underwent IVF treatment for the first time and from 1014 patients who underwent a second IVF treatment during the period 1989-1994.

The observed probabilities of achieving an ongoing pregnancy during the first and second IVF for each population are presented in Table 1.

Table 1. Number of patients and ongoing pregnancies

	Nijmegen 1991-1994			Nijmegen 1995-1996			Eindhoven			Voorburg		
	N	Pregnancies		N	Pregnancies		N	Pregnancies		N	Pregnancies	
		n	%		n	%		n	%		n	%
First IVF cycle	757	88	11.6	208	34	16.3	428	46	10.7	1424	236	16.6
Second IVF cycle	454	61	13.4	127	22	17.3	268	28	10.4	1014	149	14.7

Before starting to improve the prognostic models developed earlier on the basis of data from the Nijmegen 1991-1994 population, we studied whether the poor predictions in the Eindhoven population were due to chance. Therefore, we tested the models on data from another two populations: Nijmegen 1995-1996 and Voorburg. Moreover, as the criteria to cancel a cycle because of too many large follicles were wider in Eindhoven than in Nijmegen, we also tested the models on a selection of patients from Eindhoven after excluding any couples who discontinued the first IVF cycle because of too many follicles.

If the number of variables that can be included into a prognostic model is large in comparison with the number of ongoing pregnancies, the model will be overfitted.[1] The consequence is low predictive performance in other populations. In this study, the number of variables considered for each model was fairly large: in Model I, the patient characteristics known at the start of treatment, and in Model II, the former characteristics plus the results of the first IVF treatment cycle. To decrease the number of potential prognostic factors and thereby the possibility of overfitting, new models were developed in which only variables were allowed that showed in any of the four populations: (i) nearly complete information (i.e. less than 2% missing values) and (ii) reasonable distribution (i.e. for dichotomous variables a smallest proportion of more than 5%, on the condition that it was not known beforehand that the variable was a

strong predictor). Logistic regression analysis was used to develop a prognostic model for the Nijmegen population of 1991-1994, by employing a stepwise selection method with a p value of 0.10 and a change in c index of 0.005 as inclusion and exclusion criteria.

Next, bootstrapping was used to gain insight into the internal validity of the model and thus into the amount of overfitting of the model using the data from the Nijmegen 1991-1994 population. For bootstrapping various samples are drawn with replacements from the original population. A new prognostic model is developed for each sample and applied to the original population to estimate the c index. We used 100 samples. For each bootstrap sample, logistic regression analysis was performed using a stepwise selection method, with as sole inclusion and exclusion criterion a p value of 0.10, because the selection criterion of the change in c index could not be operated automatically in SAS. The result of bootstrapping is a nearly unbiased estimation of the optimism of the model and thus of the internal validity of the model.

Shrinkage is a calibration method that can be used to correct for overfitting. A heuristic shrinkage estimator was applied to the standardized independent variables. This shrinkage estimator is defined as (modelχ^2 - p) / modelχ^2, where modelχ^2 is the likelihood ratio [-2ln(L_1/L_2)] of the model with all the potential prognostic factors, based on the development population and p, the number of all potential prognostic factors.[9] To calculate the standardized independent variables, the mean and standard deviation were taken from the variable distribution in the development population. The result of shrinkage is that the predicted probabilities move towards the average probability. After application of the shrinkage estimator, the intercept was adjusted so that the observed and predicted numbers of pregnancies in the development population were equal. Subsequently, the models were tested again using the data from the other three populations.

If the overall probability of achieving an ongoing pregnancy in the development population is considerably different from that in the test population and this is not due to the characteristics of the population, the predictions for the test population will systematically be too high or too low. The probabilities of achieving an ongoing pregnancy in the overall test populations were used to correct for these differences in prior probabilities, so that the total predicted and observed numbers of ongoing pregnancies in the test population matched. Observed and expected numbers of women with an ongoing pregnancy in each 5% probability category, if appropriate, were compared by determining the Hosmer-Lemeshow goodness-of-fit statistic for external validation (C_v which is distributed as χ^2 with df = number of categories for the original

and shrunken model, and df = number of categories - 1 for the models which are calibrated by prior probability).[10] The effect of each of the strategies to improve validity, i.e. calibration by prior probability, shrinkage and a combination of these two methods, was evaluated on the basis of p values.

The discriminative performance of a model was evaluated by calculating the overall discrimination and assessing the predictive value of a negative test. The overall discriminative performance is expressed in the c index [i.e. (number of concordant pairs + 0.5 * number of tied pairs) / total number of pairs].[1,11] The c index can be interpreted as the probability of a correct prediction for a random pair that comprises a woman with an ongoing pregnancy and a woman without an ongoing pregnancy. The value of the c index is equal to the area under the Receiver Operating Characteristic (ROC) curve.[12] As it is of clinical importance to identify couples who have a low probability of achieving an ongoing pregnancy during IVF treatment, the emphasis is put on the discriminative performance of the models to detect couples who have a probability of achieving an ongoing pregnancy of at the most 5% or 10%. This is equivalent to the evaluation of the predictive value of a negative test with a cut-off point of 5% and 10% predicted probability, respectively. Their complements, i.e. the proportions of women who achieved an ongoing pregnancy in the group of women with a predicted probability of 0-5% or 0-10%, have been presented with mid-p exact 95% confidence intervals (CI).[13] The probability of achieving an ongoing pregnancy after one IVF treatment is at best expected to match the natural fecundity in fertile couples, which is about 30% per cycle. Therefore, the predictive value of a positive test is expected to be low and cannot be used as a criterion to determine the discriminative ability of a prognostic model.

Results

The distributions of the patient characteristics, i.e. the potential prognostic factors present at the start of the first treatment to be used in Model I, in each of the four populations are shown in Table 2. In addition to patient characteristics, the results of the first IVF treatment are potential prognostic factors in Model II. The distributions of these factors in the couples who were included in the populations for developing and testing Model II are shown in Table 3. Testing of the models developed with data from the Nijmegen 1991-1994 population with data of the other populations seemed appropriate, because the ranges of the values of the potential prognostic variables in Nijmegen 1991-1994 were comparable with those in the other populations.

Table 2 Patient characteristics in each of the four populations at the start of the first IVF cycle

	Nijmegen 1991-1994 (N=757)						Nijmegen 1995-1996 (N=208)						Eindhoven (N=428)						Voorburg (N=1424)					
	Min	Max	Mean	SD	Median	Miss (%)	Min	Max	Mean	SD	Median	Miss (%)	Min	Max	Mean	SD	Median	Miss (%)	Min	Max	Mean	SD	Median	Miss (%)
Woman's age (years)	22	47	32 9	4 0	33	0 0	22	43	33 2	4 2	33	0 0	21	43	31 9	4 1	32	0 2	22	43	32 9	4 0	33	0 0
Duration of infertility (years)	0	20 5	4 4	2 8	4 0	49 4	0 5	16	3 6	2 4	3	1 0	0	20 5	3 7	2 7	3 5	0 2	0	24 5	5 2	3 1	4 5	6 6
Basal FSH (IU/l)	<0	6 23	6 1	2 8	5 7	3 2						100 0						100 0	0 1	26 0	7 5	2 9	7 2	67 1
		n	%		Miss	(%)		n	%		Miss	(%)		n	%		Miss	(%)		n	%		Miss	(%)
≥1 Preceding gestations		256	33 8		0 0			61	29 3		0 0			160	37 4		0 0			246	17 3		0 0	
≥1 Preceding spontaneous abortions		126	16 6		0 0			33	15 9		0 0			29	8 3		18 7			97	6 8		0 5	
≥1 Preceding ectopic pregnancies		62	8 2		0 0			20	9 6		0 0			23	6 6		18 7			75	5 3		0 5	
≥1 Preceding deliveries		138	18 2		0 0			26	12 5		0 0			49	14 0		18 5			106	7 5		0 5	
Indication for IVF*																								
Tubal pathology		305	40 3		0 0			77	37 0		0 0			188	44 7		1 6			637	44 8		0 1	
Male factor		323	42 7		0 0			87	41 8		0 0			124	29 5		1 6			486	34 2		0 1	
Endometriosis		163	21 5		0 0			35	16 8		0 0			67	15 9		1 6			120	8 4		0 1	
Cervical factor		118	15 6		0 0			19	9 1		0 0			2	0 5		1 6			118	8 3		0 1	
Idiopathic infertility		133	17 6		0 0			29	13 9		0 0			74	17 6		1 6			253	17 8		0 1	
2 Ovaries		708	93 7		0 1			189	91 7		0 1			391	93 3		2 1			758	96 1		44 6	
Sperm characteristics																								
≥20*10^6/ml		630	83 2		0 0						100 0						100 0						100 0	
≥60% Normal forms		433	57 2		0 0						100 0						100 0						100 0	
≥50% Motile		416	55 0		0 0						100 0						100 0						100 0	
Quality of motility ≥4†		660	87 2		0 0						100 0						100 0						100 0	
Anti-sperm antibodies ♂ or ♀		66	8 7		0 3			0	0 0		0 0			10	2 3		0 0			11	0 8		0 0	
In sperm		38	5 0		0 3						100 0						100 0						100 0	
In woman's serum		29	3 8		0 0						100 0						100 0						100 0	

* More than one indication per couple is possible (except in the case of idiopathic infertility)

† On a scale from 1 (poorest) to 5 (best)

Table 3 Patient characteristics and results of the first IVF treatment in each of the four populations at the start of the second IVF cycle

	Nijmegen 1991-1994 (N=454)						Nijmegen 1995-1996 (N=127)						Eindhoven (N=268)						Voorburg (N=1014)					
	Min	Max	Mean	SD	Median	Miss (%)	Min	Max	Mean	SD	Median	Miss (%)	Min	Max	Mean	SD	Median	Miss (%)	Min	Max	Mean	SD	Median	Miss (%)
Woman's age (years)	24	44	33 0	3 8	33	0 0	22	44	33 3	4 2	34	0 0	21	43	31 9	4 0	32	0 0	22	43	33 1	4 0	33	0 0
Duration of infertility (years)	0 5	16 5	4 2	2 5	4	56 2	0	16	3 9	2 7	3 5	0 8	0	20 5	3 5	2 7	3 3	0 0	0	24 5	5 2	3 1	4 5	6 4
Basal FSH (IU/l)	<0 6	23	6 2	2 7	5 8	2 6						100 0						100 0	0 3	26 0	7 7	3 1	7 2	63 8
Number of follicles (>9 mm)	0	52	12 6	8 1	11	0 9	0	38	11 1	8 8	11	0 0	0	45	7 9	7 6	7	1 1	5	20	11 1	4 9	11 5	99 0
Number of oocytes	0	48	8 3	6 6	7	0 7	0	40	7 8	7 2	7	0 0	0	27	5 4	5 3	5	0 0	0	63	7 2	6 7	6	0 2
Number of fertilized oocytes	0	27	4 7	4 5	3	0 7	0	25	4 9	5 2	4	0 8	0	21	4 0	4 1	3	0 0	0	36	3 1	4 2	2	0 2
Number of transferred embryos	0	4	2 2	1 3	3	0 0	0	4	1 7	1 1	2	0 0	0	4	1 7	1 4	2	0 0	0	5	1 5	1 3	2	0 2
Number of transferred embryos of at least good quality	0	4	1 3	1 2	1	0 0	0	3	1 2	1 0	1	0 0	0	3	0 9	1 1	0	32 1	0	4	0 8	1 1	0	34 7

	n	%	Miss (%)	n	%	Miss (%)	n	%	Miss (%)	n	%	Miss (%)
≥1 Preceding gestations	155	34 1	0 0	46	36 2	0 0	106	39 6	0 0	179	17 7	0 0
≥1 Preceding spontaneous abortions	81	17 8	0 0	22	17 3	0 0	19	9 1	22 4	70	7 0	0 7
≥1 Preceding ectopic pregnancies	38	8 4	0 0	12	9 4	0 0	16	7 7	22 4	57	5 7	0 7
≥1 Preceding deliveries	83	18 3	0 0	21	16 5	0 0	24	11 5	22 4	69	6 9	0 7
Indication for IVF*												
Tubal pathology	194	42 7	0 0	43	33 9	0 0	134	50 8	1 5	463	45 7	0 0
Male factor	188	41 4	0 0	47	37 0	0 0	70	26 5	1 5	338	33 3	0 0
Endometriosis	104	22 9	0 0	24	18 9	0 0	46	17 4	1 5	87	8 6	0 0
Cervical factor	69	15 2	0 0	13	10 2	0 0	2	0 8	1 5	83	8 2	0 0
Idiopathic infertility	79	17 4	0 0	26	20 5	0 0	37	14 0	1 5	188	18 5	0 0
2 Ovaries	429	94 5	0 0	114	91 2	1 6	245	93 9	2 6	545	95 4	43 7
Sperm characteristics												
≥20*10^6 /ml	398	87 7	0 0			100 0			100 0			100 0
≥60% Normal forms	269	59 3	0 0			100 0			100 0			100 0
≥50% Motile	257	56 6	0 0			100 0			100 0			100 0
Quality of motility ≥4†	402	88 5	0 0	0		100 0			100 0			100 0
Anti-sperm antibodies, ♂ or ♀	29	6 4	0 0	0	0 0	0 0	7	2 6	0 0	9	0 9	0 0
In sperm	17	3 7	0 0			100 0			100 0			100 0
In woman's serum	12	2 6	0 0			100 0			100 0			100 0
Embryo transfer performed	383	84 4	0 0	98	77 2	0 0	178	66 4	0 0	657	64 8	0 0
Positive pregnancy test result	29	6 4	0 0	13	10 2	0 0	5	1 9	0 4	99	9 8	0 0

* More than one indication per couple is possible (except in the case of idiopathic infertility)

† On a scale from 1 (poorest) to 5 (best).

Table 4. Prognostic models for the probability of achieving an ongoing pregnancy (P) during the first IVF cycle or second IVF cycle

Type of model*:	Ln [P/(1−P)] =	SE(ß)	-2ln(L_1/L_2) df p value	Nijmegen 1991-1994 at development N	C index	C index adj.†	Shrinkage (%)
Model I.‡	− 0.3350	0.9503	14.042	757	0.612	0.584	60.64
	+ 0.8151 * ≥1 preceding gestation	0.2349	df=2				
	− 0.0620 * woman's age (years)	0.0297	p=0.0009				
Model II.§	− 1.1507	1.3907	18.068	454	0.669	0.603	49.84
	− 0.0708 * woman's age (years)	0.0372	df=3				
	− 0.7478 * idiopathic infertility	0.4528	p=0.0004				
	+ 1.8674 * embryo transfer during 1st IVF cycle	0.7328					

* For the dichotomous variables: value 1 if present, 0 if not present.

† C index adjusted for the optimism calculated by bootstrapping.

‡ Prediction at start of 1st IVF cycle regarding probability of achieving an ongoing pregnancy during 1st cycle.

§ Prediction at end of 1st IVF cycle regarding probability of achieving an ongoing pregnancy during 2nd cycle.

Table 4 describes the models developed using data from the Nijmegen 1991-1994 population. The probability of achieving an ongoing pregnancy decreased with increasing age of the woman and increased if the woman had experienced gestation before IVF treatment. The c index at development was 0.61. Table 5 presents the results of testing this model using data from the Nijmegen 1995-1996, Eindhoven and Voorburg populations. The c indexes for Model I were 0.61, 0.50 and 0.55, respectively. Only one or two women in each population had a predicted probability of 0-5%. The proportion of women with a predicted probability of 0-10% who became pregnant ranged from 9-16%. The description of Model II in Table 4 shows that the probability of achieving an ongoing pregnancy during the second IVF treatment decreased with increasing age of the woman and when idiopathic infertility was an indication for IVF, and increased if embryo transfer had been performed during the first IVF treatment. The c index of Model II at development was 0.67. After testing this model using data from the Nijmegen 1995-1996, Eindhoven and Voorburg populations, the c indexes were 0.58, 0.50 and 0.53, respectively (Table 5). The proportion of women with a predicted probability of 0-5% who achieved an ongoing pregnancy varied from 10-12%, while the proportion with a predicted probability of 0-10% who became pregnant varied from 10-13%.

Table 5 Discrimination of the original models, with or without calibration by using the prior probability and Hosmer-Lemeshow statistic of goodness-of-fit test (C_v)

Test population	N*	C index	Original									Original calibrated by prior probability
			Predicted probability 0-5%				Predicted probability 0-10%				C_v, df, p value	C_v, df, p value
			m†	n‡	%	95% CI	m	n	%	95% CI		
Model I												
Nijmegen 1995-1996	208	0 606	0	1	0 0	0 0-95 0	16	104	15 4	9 4-23 3	10 69, 4, 0 03	3 52, 3, 0 32
Eindhoven	427	0 499	0	1	0 0	0 0-95 0	14	156	9 0	5 2-14 3	5 77, 4, 0 22	13 62, 2, 0 001
Eindhoven restricted§	398	0 504	0	1	0 0	0 0-95 0	14	148	9 5	5 5-15 0	4 63, 4, 0 33	10 27, 3, 0 02
Voorburg	1424	0 545	0	2	0 0	0 0-77 6	131	829	15 8	13 4-18 4	87 28, 4, 0 0001	19 20, 6, 0 004
Model II												
Nijmegen 1995-1996	127	0 579	3	30	10 0	2 6-24 9	5	47	10 6	4 0-22 0	7 77, 3, 0 05	3 35, 4, 0 50
Eindhoven	264	0 503	9	86	10 5	5 2-18 3	10	104	9 6	5 0-16 5	14 77, 4, 0 005	10 81, 3, 0 01
Eindhoven restricted§	238	0 610	2	61	3 3	0 6-10 4	3	78	3 8	1 0-10 1	4 86, 4, 0 30	0 16, 2, 0 92
Voorburg	1014	0 527	44	354	12 4	9 3-16 2	59	440	13 4	10 5-16 8	153 89, 6, 0 0001	110 80, 6, 0 0001

* Patients with missing values on one or more of the variables were excluded, i e for Eindhoven total and restricted 1 for Model I, 4 for Model II

† Number of women who achieved an ongoing pregnancy

‡ Total number of women

§ Excluding 29 and 26 women whose first IVF treatment was cancelled because of too many follicles present for Models I and II, respectively.

As the Eindhoven centre cancelled IVF treatment on the basis case of too many large follicles more readily than the Nijmegen centre, 29 couples whose first IVF treatment had been cancelled because of too many follicles (26 of them had a second IVF treatment) were excluded from the Eindhoven test population. Forty-six out of the remaining 399 women (11.5%) achieved an ongoing pregnancy during the first IVF cycle, while 21 out of the 242 women (8.7%) became pregnant during the second IVF cycle. Testing the models using data from this restricted Eindhoven population resulted in c indexes of 0.50 for Model I and 0.61 for Model II. The proportion of women with a predicted probability of 0-10% who became pregnant was 10% for Model I (Table 5). For Model II, one third of the couples had a predicted probability of 0-10% and the observed proportion of women who achieved an ongoing pregnancy in this group was 4% (Table 7).

It was also studied whether cancellation formed an explanation for the poor predictive performance of the models on the Voorburg population. In Voorburg, however, IVF cycles were never cancelled because too many follicles were present. When we only considered the women who had undergone embryo transfer during the first IVF cycle, the probability of achieving an ongoing pregnancy during the first IVF treatment could be predicted accurately at the moment of embryo transfer with a model described in our previous study.[8] The c index was 0.84 and the proportion of women with a predicted probability of 0-5% that became pregnant was 1.7% (95% CI 0.8-3.0%). Such a low predicted probability occurred in 60% of the patients for whom information was available about prognostic factors: prior gestation and numbers of fertilized oocytes, embryos transferred and good quality embryos transferred. The proportions of first IVF cycles in which no oocyte aspiration was performed were comparable between Voorburg and Nijmegen 1991-1994 [9.6% (137 out of the 1424 women) and 7.4% (56 out of the 757 women), respectively]. In Nijmegen and Voorburg, this part of the IVF treatment, between the start of the first IVF cycle and embryo transfer, differed only in the ovulation stimulation protocol: in Nijmegen a long protocol of gonadotrophin-releasing hormone agonist was used, whereas in Voorburg, in general, no agonist was used. This resulted more often in premature luteinizing hormone (LH) peaks in women in Voorburg than in Nijmegen (absolute numbers are unknown), with as consequence less follicles available for aspiration because of ovulations. However, when only women without a premature LH peak were considered at the Voorburg centre, both prognostic models did not predict any better. Thus the differences in type of stimulation protocols seemed not to be an explanation for the poor performance of the models in Voorburg.

Table 6. Shrinkage of the prognostic models for the probability of achieving an ongoing pregnancy (P) during the first IVF cycle or second IVF cycle

Type of model*:	Ln [P/(1-P)] =	SE(ß)	$-2\ln(L_1/L_2)$ df p value	Nijmegen 1991-1994 at development	
				N	C index
Model I.†	– 2.0533	0.1200	14.042	757	0.611
	+ 0.3859 * 0.6064 * (≥1 preceding gestation – 0.33818) / 0.47340	0.1112	df=2 p=0.0009		
	– 0.2458 * 0.6064 * [woman's age (years) – 32.8666] / 3.96605	0.1177			
Model II.‡	– 1.9138	0.1700	18.068	454	0.669
	– 0.2698 * 0.4984 * [woman's age (years) – 33.0419] / 3.8083	0.1415	df=3 p=0.0004		
	– 0.2838 * 0.4984 * (idiopathic infertility – 0.17401) / 0.37954	0.1719			
	+ 0.6790 * 0.4984 * (embryo transfer during 1st IVF cycle – 0.84361) / 0.36362	0.2665			

* For the dichotomous variables: value 1 if present, 0 if not present.

† Prediction at start of 1st IVF cycle regarding probability of achieving an ongoing pregnancy during 1st cycle.

‡ Prediction at end of 1st IVF cycle regarding probability of achieving an ongoing pregnancy during 2nd cycle.

Restriction of model development to variables with less than 2% missing values and at least a 5% proportion of dichotomous variables in any of the four populations (see Tables 2 and 3) led to the exclusion of the following potential prognostic factors for Model I: the duration of infertility, basal FSH, preceding spontaneous abortions, ectopic pregnancies and deliveries, cervical factor, number of ovaries, sperm characteristics and anti-sperm antibodies in the male or female. In addition, the following were excluded for Model II: the number of follicles, the number of transferred embryos of at least good quality and a positive pregnancy test result. Despite these restrictions, the models developed now were exactly the same as the original ones (see Table 4). Accordingly, the test results presented in Table 5 applied to this first attempt to increase internal validity.

Bootstrapping, using the variables selected above, showed an optimism of 0.03 for Model I and 0.07 for Model II. Thus the c indexes adjusted for this amount of overfitting were 0.61 - 0.03 = 0.58 and 0.67 - 0.07 = 0.60 for Models I and II, respectively (Table 4).

The shrinkage estimate was 61% [(15.24 - 6) / 15.24] for Model I and 50% [(19.94 - 10) / 19.94] for Model II. When the shrinkage estimator was incorporated

Table 7 Discrimination of the models after shrinkage, with or without calibration by using the prior probability and Hosmer-Lemeshow statistic of goodness-of-fit test (C_v)

Test population	N*	C index	Shrinkage									Shrinkage, calibrated by prior probability
			Predicted probability 0-5%				Predicted probability 0-10%				C_v, df, p value	C_v, df, p value
			m†	n‡	%	95% CI	m	n	%	95% CI		
Model I												
Nijmegen 1995-1996	208	0 606	0	0	-	-	7	79	8 9	4 0-16 7	8 77, 3, 0 03	0 60, 2, 0 74
Eindhoven	427	0 499	0	0	-	-	10	100	10 0	5 2-17 1	3 41, 3, 0 33	5 54, 2, 0 06
Eindhoven restricted§	398	0 504	0	0	-	-	10	99	10 1	5 2-17 3	2 81, 3, 0 42	2 15, 2, 0 34
Voorburg	1424	0 545	0	0	-	-	91	611	14 9	12 2-17 9	52 07, 3, 0 0001	5 10, 3, 0 16
Model II												
Nijmegen 1995-1996	127	0 565	1	4	25 0	1 3-75 8	5	37	13 5	5 1-27 4	4 28, 2, 0 12	0 73, 2, 0 69
Eindhoven	264	0 556	1	13	7 7	5 0-17 6	9	93	9 7	4 6-15 2	4 19, 3, 0 24	4 65, 2, 0 10
Eindhoven restricted§	238	0 646	1	11	9 1	0 5-37 3	2	67	3 0	0 5-9 5	4 35, 3, 0 23	0 42, 1, 0 52
Voorburg	1014	0 558	8	61	13 1	6 3-23 4	50	387	12 9	9 8-16 5	27 82, 3, 0 0001	19 08, 3, 0 0003

* Patients with missing values on one or more of the variables were excluded, i e for Eindhoven total and restricted 1 for Model I, 4 for Model II

† Number of women who achieved an ongoing pregnancy

‡ Total number of women

§ Excluding 29 and 26 women whose first IVF treatment was cancelled because of too many follicles present for Models I and II, respectively.

into Model I (described in Table 6) and tested on the three populations, the c indexes remained the same as before shrinkage (Table 7). None of the women had a predicted probability of 0-5%. The proportion of women with a predicted probability of 0-10% who became pregnant was 9-15%. For Model II, the c indexes were 0.57, 0.56 and 0.56 for Nijmegen 1995-1996, Eindhoven en Voorburg, respectively. The proportion of women with a predicted probability of 0-5% who became pregnant was between 8% and 25%, whereas in the women with a predicted probability of 0-10%, between 10% and 14% became pregnant. For the restricted Eindhoven population that excluded the women whose first IVF cycle had been cancelled because of too many follicles, the c index was 0.50. The proportion of women with a predicted probability of 0-10% who became pregnant was 10% after shrinkage of Model I; after shrinkage of Model II, these values were 0.65 and 3%, respectively (Table 7). Calibration with the prior overall probability did not consistently lead to a better fit of the models (see goodness-of-fit statistic results in Tables 5 and 7).

Discussion

This study showed that the models developed on the basis of the data from Nijmegen 1991-1994 did not show a good predictive performance in the three test populations. As expected, the overall discriminative performances of the models were the highest in the Nijmegen 1995-1996 population, but the predictions of a 0-5% or 0-10% probability of becoming pregnant were imprecise because of too small numbers of women with such a low predicted probability. There was one exception: when the women whose first IVF treatment had been cancelled because of too many follicles were excluded from the Eindhoven population, the prognostic model for ongoing pregnancy after the second treatment predicted well which patients had a low probability of achieving an ongoing pregnancy. In this population of 238 couples, a predicted probability of less than 10% occurred in those (i) without embryo transfer during the first IVF cycle, irrespective of the woman's age or indication for IVF (n = 64), or (ii) with embryo transfer and idiopathic infertility who were older than 30 years at the first treatment (n = 12), or (iii) with embryo transfer and no idiopathic infertility who were older than 41 years at the first treatment (n = 2). In the Voorburg population, we did not succeed in finding a subsample in whom the predictive performance was good. Perhaps the predictive performance of the models at the other centres was hampered because of differences in the definitions of idiopathic infertility between the centres. In the case of idiopathic infertility, patients were offered IVF treatment after infertility duration of at least 3 years in Nijmegen and Eindhoven,

whereas in Voorburg it was only offered to patients after 4 years of infertility. Consensus in definitions and treatment protocols would facilitate the creation of externally valid prognostic models for ongoing pregnancy after IVF treatment. Moreover, it is possible that there are factors which were incorrectly not identified to be prognostic, because information was incomplete or absent. This might apply to, for instance, basal FSH level and infertility duration.

For bootstrapping, variable selection was only based on the p value, whereas for the original model, the selection was based on both the p value and the change in c index. In comparison with the original models (Table 4), selection based on only the p value gave the same results for Model I, but for Model II, the woman's age was excluded, while the number of fertilized oocytes was included. The c index of the original model was slightly higher than that of the model developed for bootstrapping (0.67 and 0.66, respectively). The amount of variation explained by the original model was comparable [$-2\ln(L_1/L_2) = 18.07$ df = 3 p = 0.0004 and $-2\ln(L_1/L_2) = 16.71$ df = 3 p = 0.0008, respectively]. Thus, because of the less restrictive selection criterion, the optimism calculated for both models might have been somewhat overestimated and thus the internal validity underestimated. Nevertheless, bootstrapping revealed that the models were overfitted and low c indexes were found for the original population after correction for the optimism. This means that the overall predictive performance of these models in the test populations would not be high. However, it does not preclude a reasonable discriminative performance at the lower probability range.

In concordance with the results of bootstrapping, the shrinkage estimators for Model I and II were fairly high: 61% and 50%, respectively. Although the overall predictive performances (c indexes) in the test populations improved in some instances, the predictive values of the negative tests did not increase noticeably (Tables 5 and 7).

The presumed benefit of calibration using the prior probability was doubtful, even in this ideal situation in which the exact 'prior' probability was known. This seems to indicate that there was no centre- or time-period-specific probability, but only differences in the overall probability because of another distribution of population characteristics.

In conclusion, Model I designed to predict at the start of the first IVF treatment the probability of achieving an ongoing pregnancy during the first IVF treatment, showed no discriminative ability as the c indexes were fairly low, the number of couples with a predicted probability of 0-5% was negligible and the predictions of a probability of 0-10% imprecise. Model II, designed to predict at the end of the first IVF treatment the probability of achieving an ongoing pregnancy during the second IVF treatment, only

could identify well women with a low probability of achieving an ongoing pregnancy in the Eindhoven population after excluding the women whose first IVF treatment had been cancelled because of too many follicles. Despite this finding, the external validity of the presented prognostic models can still be questioned. The experience gained in this search for prognostic models for IVF, emphasizes the need for caution when using such models without proper testing. The models presented here require further development and testing. In particular, additional testing of the discriminative performance of the models on future patients at the Nijmegen centre is warranted, because they have not yet been adequately evaluated as the number of patients was too small. Shrinkage might be of some use to correct for overfitting. The discriminative performance of a prognostic model should be evaluated by focusing on the predictive value of a test in the low range of probability (the negative test) rather than by parameters reflecting the overall discrimination, such as the c index.

References

1. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-387.
2. Guzick DS, Balmaceda JP, Ord T, Asch RH. The importance of egg and sperm factors in predicting the likelihood of pregnancy from gamete intrafallopian transfer. Fertil Steril 1989;52:795-800.
3. Hughes EG, King C, Wood EC. A prospective study of prognostic factors in in vitro fertilization and embryo transfer. Fertil Steril 1989;51:838-844.
4. Haan G, Bernardus RE, Hollanders JMG, Leerentveld RA, Prak FM, Naaktgeboren N. Results of IVF from a prospective multicentre study. Hum Reprod 1991;6:805-810.
5. Nelson JR, Huppert L, Corson SL, *et al.* Predicting success of gamete intrafallopian transfer. Fertil Steril 1993;60:116-122.
6. Eimers JM, Te Velde ER, Gerritse R, Vogelzang ET, Looman CWN, Habbema JDF. The prediction of the chance to conceive in subfertile couples. Fertil Steril 1994;61:44-52.
7. Collins JA, Burrows EA, Willan AR. The prognosis for live birth among untreated infertile couples. Fertil Steril 1995;64:22-28.
8. Stolwijk AM, Zielhuis GA, Hamilton CJCM, Straatman H, Hollanders JMG, Goverde HJM, Van Dop PA, Verbeek ALM. Prognostic models for the probability of achieving an ongoing pregnancy after in vitro fertilization and the importance of testing their predictive value. Hum Reprod 1996;11:2298-2303.
9. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. Stat Med 1990;9:1303-1325.
10. Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley & Sons, Inc. 1989:171-173.
11. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247:2543-2646.
12. Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology 1982;143:29-36.
13. Vollset SE. Confidence intervals for a binomial proportion. Stat Med 1993;12:809-824.

3.4. The impact of the woman's age on the success of standard and donor in vitro fertilization

A.M. Stolwijk, G.A. Zielhuis, M.V. Sauer, C.J.C.M. Hamilton, R.J. Paulson

Objective: To study the effect of the age of the woman who provides the oocytes or who receives the embryos on results of in vitro fertilization-embryo transfer (IVF-ET). Design: Historical cohort study. Multivariate regression analysis was used to study the age effect continuously and after adjustment for confounding. Setting: Patients of the University of Southern California, Los Angeles, California. Patients: Couples who underwent standard (N = 277) or donor IVF-ET (N = 294) between January 1991 and July 1995. Interventions: One cycle of standard or donor IVF-ET. Main outcome measures: Successive IVF outcomes from number of oocytes to ongoing pregnancy and several measures of pregnancy loss. Results: The number of oocytes decreased with aging of the oocyte provider. More women who received oocytes from donors aged 20 to 23 years had at least one good embryo transferred than women who received oocytes from older donors. The age of the woman who received the embryos had no effect on IVF outcomes. In women older than 40 years who underwent standard IVF, the probability of pregnancy decreased. No such relationships were observed for donor IVF, but all the oocyte donors were younger. Conclusions: An age effect for ongoing pregnancy was only found in women older than 40 years who underwent standard IVF independent of the lower number of oocytes and suggests decreasing oocyte quality.

Fertility and Sterility (in press)

Introduction

Diminished fecundity with increasing woman's age is well-documented. An increase in anovulatory cycles has been found in women after age 40 years[1] and in perimenopausal women who experienced a sudden break in menstrual cyclicity after years of regular cycles.[2] Pregnancy rates are known to decline with increasing woman's age after artificial insemination with donor semen,[3-5] after standard in vitro fertilization-embryo transfer treatment[6-10] and after standard IVF-ET using intracytoplasmatic sperm injection (ICSI).[11] Nevertheless it is still somewhat controversial whether the woman's age during IVF treatment has an effect on oocyte production, oocyte quality or uterine receptivity. Treatment with donor oocytes offers a means of studying these effects separately. Sauer *et al*,[12] Abdalla *et al*,[13] Navot *et al*[14] and Check *et al*[15] found an effect of the woman's age on the quality of the oocyte, but no or only a negligible effect on uterine receptivity. In more recent studies, Sauer and coworkers[16,17] have confirmed that there is no aging effect of the uterus in women undergoing donor IVF. However, Levran *et al*,[18] Flamigni *et al*,[19] Borini *et al*[20] and Cano *et al*[21] found an age effect also on uterine receptivity. In the former three studies, the conclusion was based on decreased implantation or pregnancy rates in older recipients, whereas Cano *et al*[21] based their conclusion on increased abortion rates.

Meldrum[22] suggested that the differences in hormone replacement protocols may explain why several studies found an aging effect of the uterus while others did not. He observed that a dose of 100 mg progesterone (P) per day administered intramuscularly (IM) could fully replenish the uterine capacity. This effect was corroborated by Check *et al*,[15] who could not find an aging effect of the uterus when hormone supplementation was prolonged until the endometrium thickness was ≥10 mm. However, this factor cannot explain the aging effects of the uterus found by Flamigni *et al*,[19] Cano *et al*[21] and Borini *et al*,[20] whose patients were all supplemented with high doses of P. Moreover, Meldrum[22] suggested that the low rate of spontaneous abortions observed in older women after donor IVF could be explained by the corrective effect of the hormone replacement. However, Cano *et al*[21] employed high doses of P and found an increased rate of spontaneous abortion in women of ≥40 years.

Other explanations for the controversy in results between studies might be found in the characteristics of the population and the methods used for analysis. Most of the studies that used data from donor IVF[12,14,15,17,18,20-22] only examined differences in IVF results between two age groups. Existing age effects might have been missed because the chosen cut-off point did not maximize the differences between the two age groups. Also the range in age might have been too small to detect an effect. Moreover, often no adjustment was

made for the confounding effects of factors such as type of infertility and man's age.

In this study we addressed the question of whether there is an effect of the woman's age on IVF outcomes by using a data base containing a wide age range of women who provided oocytes and women who received the embryos. Multivariate regression analysis was used to fully account for the age of the oocyte providers and recipients and to fully account for any potential confounding effects.

Materials and Methods

To study the effect of the woman's age on IVF results, data were used from patients who underwent standard or donor IVF at the University of Southern California, Los Angeles, California, between January 1991 and July 1995. Before acceptance for donor IVF, the women were screened to ensure that they were in good health.[23] A normal uterine cavity was documented in all patients by a normal hysterosalpingogram or prior hysteroscopy. Additionally, all recipients underwent a 'mock cycle' of steroid replacement with an endometrial biopsy, indicating appropriate histologic response to the regimen. The protocol for ovarian hyperstimulation in the oocyte donor women was the same as that in the women who underwent standard IVF. Most women received a long protocol of gonadotropin releasing hormone (GnRH) agonist, started at the midluteal phase of the previous cycle, and human menopausal gonadotropin. Hormone supplementation of the oocyte recipients consisted of oral oestradiol (E_2), starting approximately 3 days before the donor's initiation of hMG treatment. Progesterone supplementation was started on the morning after human chorionic gonadotropin had been administered to the donor. On the first day of P supplementation, the recipient received 50 mg of P IM followed by 100 mg daily. Pregnancies were supported by E_2 and 100 mg P daily until 100 days after embryo transfer. The women who underwent standard IVF received 25 mg P IM daily commencing on the day of embryo transfer and continuing until approximately 9 weeks of gestation. These protocols have been described more extensively by Paulson *et al.*[24]

Only the data from IVF cycles utilizing controlled ovarian hyperstimulation were included. We excluded any cycles in which ICSI had been performed and all those in which there was no male partner. Data from women who underwent standard IVF were only used if follicle aspiration had been performed. To prevent bias because of repeated cycles after unsuccessful treatment, only the first IVF treatment cycle of each woman was analyzed. If donor IVF had been performed after standard IVF, only the donor IVF cycle was included in the analyses. Information on cryopreserved embryo transfer was disregarded. A total of 571 couples met the criteria for inclusion: 277 women underwent standard IVF and 294 donor IVF. All the oocytes retrieved after one controlled ovarian

hyperstimulation cycle in a donor woman were donated to only one recipient. Each donor woman could give more than one donation. None of the oocyte donors underwent IVF treatment themselves. Characteristics of the couples are given in Table 1; the age distribution is shown in Figure 1.

Table 1. Characteristics of the couples who underwent standard IVF or donor IVF

Characteristic		Standard IVF (N = 277)	Donor IVF (N = 294)
$Age_{provider}$ (years)	Range	24-45	20-37
	Median	36	30
	Mean	35.4 (4.5)	29.2 (4.0)
	(SD)	0	34
	Missing		
$Age_{recipient}$ (years)	Range	24-45	24-59
	Median	36	43
	Mean	35.4 (4.9)	42.2 (6.2)
	(SD)	0	0
	Missing		
Man's age (years)	Range	23-66	24-74
	Median	38	42
	Mean	37.5 (6.3)	42.2 (7.9)
	(SD)	22	20
	Missing		
Infertility factor woman:			
Tubal pathology	n (%)	126 (49.8)	0 (0.0)
Endometriosis	n (%)	27 (10.7)	0 (0.0)
Transitional menopause*	n (%)	0 (0.0)	160 (55.7)
Menopause†	n (%)	0 (0.0)	111 (38.7)
Genetic	n (%)	0 (0.0)	9 (3.1)
Chemotherapy	n (%)	0 (0.0)	7 (2.4)
None‡	n (%)	100 (39.5)	0 (0.0)
	Missing	24	7
Male infertility factor	n (%)	90 (34.2)	113 (40.7)
	Missing	14	14
Use of donor sperm	n (%)	25 (9.1)	33 (11.2)
	Missing	2	0
Long GnRH-agonist ovarian stimulation protocol	n (%)	237 (89.4)	271 (95.4)
	Missing	12	10

* Including previous IVF failure and elevated basal FSH levels.

† Including premature ovarian failure and castration.

‡ Including unexplained infertility (n = 46) and male factor infertility only (n = 49); (for 5 couples missing information on male factor).

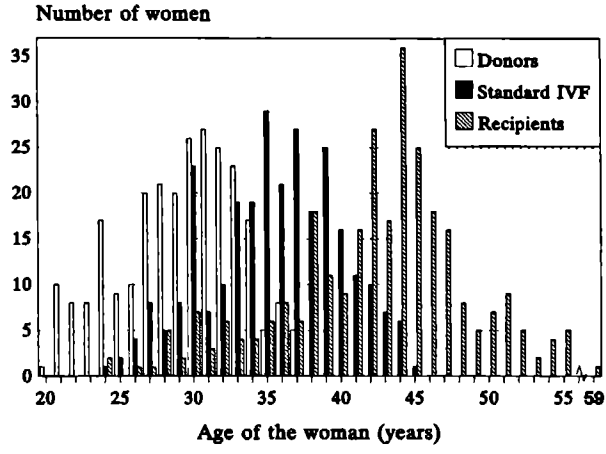


Figure 1. Age distribution of the oocyte donors, oocyte recipients and women who underwent standard IVF

The following outcomes were recorded from each aspiration cycle: number of oocytes retrieved, number of fertilized oocytes (counted 1 day after insemination), number of transferred embryos of good quality (at 48 hours after insemination: even blastomeres and <10% fragmentation if <6 blastomeres or <25% fragmentation if ≥6 blastomeres), positive pregnancy test result, clinical pregnancy (≥1 gestational sac observed by ultrasound), and ongoing pregnancy (>12 weeks after embryo transfer). In addition, the occurrence of no implantation after embryo transfer, pregnancy loss (within 12 weeks after embryo transfer) after a positive pregnancy test result, and spontaneous abortion (within 12 weeks after embryo transfer) after a clinical pregnancy were studied in relation to the woman's age.

The impact of age was studied in three phases. Phase I: the relationships with the age of the woman who provided the oocytes ($age_{provider}$) and the number of oocytes, fertilized oocytes and number of good quality embryos transferred were studied for standard and donor IVF separately. Phase II: the effect of the age of the woman who received the embryos ($age_{recipient}$) on IVF outcomes after embryo transfer was studied. Owing to the fact that each woman who underwent standard IVF was both the oocyte provider and embryo receiver, it could not be determined whether a correlation between $age_{recipient}$ and implantation or pregnancy outcome was due to oocyte or uterus factors. Therefore, the associations between $age_{recipient}$ and implantation and pregnancy outcome data were examined only in donor IVF. Phase III: if such an association with $age_{recipient}$ exists, the relationships of $age_{provider}$ with treatment outcomes after embryo transfer can only be studied in women who undergo donor IVF. If there is no relationship with $age_{recipient}$, it can be assumed that there is no aging effect of the uterus in standard IVF as well. In that case,

information from both standard and donor IVF can be used to investigate whether there remains an effect of the $age_{provider}$ on the outcomes of IVF after embryo transfer.

Linear regression was used if the dependent variable had a continuous distribution, whereas logistic regression analysis was used if the dependent variable was dichotomous. Results of logistic regression analysis are presented in the form of odds ratios (OR) with 95% confidence intervals (CI). To study whether the relationship between the IVF result and the woman's age showed a change in trend, categories were made for each 5 years of age if the number of couples in the age group was reasonably large. As a reference category, the age group with the largest number of couples was chosen. When a change in age trend was suspected on the basis of the regression coefficients of these categories, we performed several regression analysis to locate the best cut-off point, which allowed for a change in the continuous age effect. The best cut-off point was selected by choosing the model with the highest explained variation in IVF result by age as expressed by the F-test or the likelihood ratio test $[-2\ln(L_1/L_2)]$.

An association between the woman's age and the outcome of IVF can be confounded by the effects of several factors. Confounding can be detected by observing a substantial change in the regression coefficients of the age variables after adjustment for confounding. To adjust for confounding, all the potential confounders were included into the linear or logistic regression model as dummy variables. For the number of oocytes retrieved, potential confounding factors were the type of hormonal ovulation stimulation (long protocol of GnRH-agonist versus others) and, only for standard IVF, type of infertility of the woman who provided the oocytes (tubal pathology, endometriosis or none). For the relationships with fertilized oocytes and the number of good quality embryos transferred, potential confounders included also a male infertility factor and the age of the male partner. For the other IVF results, additional potential confounders were the type of infertility of the embryo recipient (for standard IVF: tubal pathology, endometriosis or none; for donor IVF: transitional menopause, menopause, chemotherapy or genetic). For a description of the definitions of the types of infertility in women see Sauer *et al.*[16] Male infertility factor was defined as abnormal semen analysis, vasectomy reversal, failed hamster test or prior failed fertilization. If donor sperm was used, a male infertility factor was considered to be absent. Dummy variables of 5-year groups were constructed for male age, which provided the possibility of including a dummy variable for male age if donor sperm was used.

Table 2. IVF outcomes of the couples who underwent standard IVF or donor IVF

IVF outcome		Standard IVF (N = 277)	Donor IVF (N = 294)
No. of Oocytes /aspiration	Range	0-57	0-58
	Median	12	15
	Mean (SD)	13.1 (8.2)	16.4 (7.8)
	Missing	0	0
No. of Fertilized oocytes /aspiration	Range	0-31	0-24
	Median	5	7
	Mean (SD)	6.3 (4.9)	8.0 (4.8)
	Missing	0	0
No. of Transferred embryos /aspiration*	Range	0-8	0-9
	Median	5	5
	Mean (SD)	3.8 (1.8)	4.3 (1.5)
	Missing	0	0
No. of Good embryos transferred /aspiration*	Range	0-5	0-5
	Median	0	0
	Mean (SD)	0.6 (1.0)	0.5 (0.8)
	Missing	0	0
No. of Embryos implanted /aspiration*	Range	0-3	0-5
	Median	0	0
	Mean (SD)	0.4 (0.8)	0.6 (1.0)
	Missing	14	5
% Embryos implanted /embryo transfer†	Range	0-100	0-100
	Median	0	0
	Mean (SD)	9.1 (18.2)	13.4 (21.9)
	Missing	14	5
≥1 Good embryo transferred /aspiration*	n (%)	91 (33.2)	90 (30.6)
	Missing	0	0
Positive pregnancy test result /aspiration*	n (%)	83 (30.9)	109 (37.2)
	Missing	5	1
Clinical pregnancy /aspiration*	n (%)	70 (26.0)	102 (34.8)
	Missing	5	1
Ongoing pregnancy /aspiration*	n (%)	62 (23.0)	81 (27.6)
	Missing	5	1
No implantation /embryo transfer†	n (%)	117 (74.4)	175 (64.1)
	Missing	14	5
Pregnancy loss /positive pregnancy test result‡	n (%)	21 (25.3)	28 (25.7)
	Missing	5	1
Spontaneous abortion /clinical pregnancy§	n (%)	8 (11.4)	21 (20.6)
	Missing	5	1

* Excluding 3 women with all embryos frozen after standard IVF.

† Excluding 25 and 16 women without embryo transfer, respectively, for standard and donor IVF.

‡ Excluding 189 and 184 women without a positive pregnancy test result, respectively, for standard and donor IVF.

§ Excluding 202 and 191 women without a clinical pregnancy, respectively, for standard and donor IVF.

Results

The IVF results of the couples who underwent standard or donor IVF are shown in Table 2. No oocytes were retrieved after follicle aspiration in one woman who underwent standard IVF and in one donor woman. No fertilization occurred in another 19 couples with standard IVF and in 13 couples with donor IVF. Additionally no embryos were transferred in another five women with standard IVF and two women with donor IVF. In three women with standard IVF, all the embryos were frozen.

The results of the effects of $age_{provider}$ and $age_{recipient}$ are shown in Tables 3 and 4A-C. As confounding was present in many instances, only the adjusted results are shown for the analyses with 5-year age groups. If a change in age effect by using a cut-off point explained the variation in IVF outcomes better than a continuous age effect (as presented in Tables 3 and 4A-C), the regression coefficients or odds ratios accompanying the lines before and after the cut-off point are presented in the text. They are not shown in Tables 3 and 4A-C. To calculate percentages of women who had a specific IVF outcome, the most common distribution of confounders was assumed: tubal factor for standard IVF and transitional menopause for donor IVF, a long protocol of GnRH-agonist, no male infertility factor, no donor sperm and male age of 38 years for standard IVF and 42 years for donor IVF.

Table 3. Regression coefficients (ß) of linear regression analysis for crude and adjusted associations* between the woman's age from whom the oocytes are retrieved (=$age_{provider}$) and to whom the embryos's are transferred (=$age_{recipient}$) and IVF outcomes for standard and donor IVF separately

IVF outcome		A Standard IVF and age (N = 277)†					B Donor IVF and $age_{provider}$ (N = 260)				
		Crude	Adj	In categories (years) Adjusted‡§			Crude	Adj	In categories (years) Adjusted‡‖		
				24-29	30-34	40-45			20-24	25-29	35-37
No of oocytes retrieved /aspiration	ß	-0 51	-0 47	6 25	0 79	-1 11	-0 29	-0 24	2 16	0 83	-0 86
	SE(ß)	0 11	0 11	1 74	1 20	1 45	0 12	0 12	1 37	1 14	2 00
	n	277	245		245		260	253		253	
	P	0 0001	0 0001		0 002		0 01	0 05		0 38	
No of oocytes fertilized /aspiration	ß	-0 16	-0 04	1 50	-0 37	1 14	-0 14	-0 04	1 40	0 04	3 27
	SE(ß)	0 07	0 08	1 17	0 78	0 97	0 07	0 08	0 88	0 74	1 47
	n	277	225		225		260	227		227	
	P	0 01	0 64		0 27		0 06	0 61		0 07	

* Variables adjusted for were, for 'number of oocytes retrieved' type of hormonal ovulation stimulation and, only for standard IVF, type of infertility of the woman who provided the oocytes, for 'fertilized oocytes' additionally male infertility factor and man's age

† For standard IVF $age_{provider} = age_{recipient}$

‡ P values in this column refer to the variation in IVF result explained by age as expressed in the F-test, n refers to the sample size

§ Reference category 35-39 years

‖ Reference category 30-34 years

Table 4A. Odds ratios (OR) and 95% confidence intervals (CI) for crude and adjusted associations* between the woman's age and IVF outcomes for standard IVF

IVF outcome		A. Standard IVF and age (N = 277)¶				
		Crude	Adjusted	In categories (years) Adjusted** ††		
				24-29	30-34	40-45
At least 1 good	OR	0.98	1.06	1.16	0.82	2.19
embryo transferred	95% CI	0.93-1.04	0.97-1.15	0.40-3.38	0.39-1.71	0.82-5.86
/aspiration†	n	274	222		222	
	P	0.54	0.18		0.33	
Positive pregnancy	OR	0.99	0.97	0.52	1.55	0.55
test result	95% CI	0.93-1.05	0.89-1.06	0.14-1.90	0.75-3.19	0.19-1.60
/aspiration†	n	269	217		217	
	P	0.63			0.15	
Clinical pregnancy	OR	0.96	0.95	0.47	1.49	0.47
/aspiration†	95% CI	0.91-1.03	0.88-1.04	0.12-1.87	0.71-3.11	0.15-1.44
	n	269	217		217	
	P	0.25	0.29		0.11	
Ongoing pregnancy	OR	0.96	0.96	0.56	1.70	0.43
/aspiration†	95% CI	0.90-1.02	0.87-1.05	0.13-2.30	0.78-3.68	0.13-1.42
	n	269	217		217	
	P	0.18	0.32		0.09	
No implantation	OR	1.07	1.09	1.41	0.47	2.17
/embryo transfer‡	95% CI	1.00-1.14	1.00-1.20	0.33-6.04	0.21-1.05	0.68-6.96
	n	238	196		196	
	P	0.06	0.06		0.05	
Pregnancy loss	OR	1.11	1.07	0.74	0.70	2.71
/positive pregnancy	95% CI	0.97-1.27	0.89-1.28	0.05-10.17	0.17-2.93	0.31-23.79
test§	n	76	69		69	
	P	0.12	0.46		0.75	
Spontaneous abortion	OR	1.08	1.01		0.48‡‡	
/clinical pregnancy‖	95% CI	0.88-1.32	0.78-1.30		0.07-3.28	
	n	65	59		59	
	P	0.47	0.97		0.45	

* Variables adjusted for were, for 'number of good quality embryos transferred': type of hormonal ovulation stimulation, male infertility factor, man's age, and, only for standard IVF, type of infertility of the woman who provided the oocytes, for the 'other IVF results': additionally type of infertility of the embryo recipient.

† Excluding 3 women with all embryos frozen.

‡ Excluding 25 women without embryo transfer.

§ Excluding 189 women without a positive pregnancy test result.

‖ Excluding 202 women without a clinical pregnancy.

¶ For standard IVF $age_{provider} = age_{recipient}$.

** P values in this column refer to the variation in IVF result explained by age as expressed in the likelihood ratio test, n refers to the sample size.

†† Reference category: 35-39 years (OR = 1).

‡‡ Age group 20-34 versus ≥35 years.

Table 4B. Odds ratios (OR) and 95% confidence intervals (CI) for crude and adjusted associations* between the woman's age from whom the oocytes are retrieved (= $age_{provider}$) and IVF outcomes for donor IVF

IVF outcome		B. Donor IVF and $age_{provider}$ (N = 260)				
		Crude	Adjusted	In categories (years) Adjusted‖¶		
				20-24	25-29	35-37
At least 1 good	OR	0.95	0.94	1.98	0.97	1.37
embryo transferred	95% CI	0.89-1.02	0.87-1.01	0.91-4.31	0.48-1.94	0.36-5.17
/aspiration	n	260	227		227	
	P	0.14	0.08		0.31	
Positive pregnancy	OR	0.99	0.98	1.82	0.88	1.90
test result	95% CI	0.93-1.06	0.91-1.05	0.83-3.99	0.45-1.73	0.52-6.91
/aspiration	n	260	227		227	
	P	0.78	0.50		0.27	
Clinical pregnancy	OR	0.99	0.98	1.79	0.90	2.49
/aspiration	95% CI	0.93-1.05	0.91-1.06	0.81-3.96	0.45-1.79	0.67-9.29
	n	260	227		227	
	P	0.66	0.63		0.22	
Ongoing pregnancy	OR	1.00	1.00	1.21	0.93	2.21
/aspiration	95% CI	0.94-1.07	0.92-1.08	0.51-2.89	0.44-1.96	0.57-8.47
	n	260	227		227	
	P	0.97	0.95		0.65	
No implantation	OR	1.00	1.01	0.61	1.19	0.46
/embryo transfer†	95% CI	0.94-1.07	0.94-1.09	0.27-1.35	0.59-2.42	0.12-1.70
	n	244	215		215	
	P	0.89	0.73		0.29	
Pregnancy loss	OR	0.97	0.99	1.81	0.64	1.47
/positive pregnancy	95% CI	0.87-1.09	0.87-1.13	0.47-7.03	0.15-2.75	0.11-19.47
test‡	n	86	76		76	
	P	0.61	0.92		0.64	
Spontaneous abortion	OR	0.96	0.99	2.11	0.87	1.88
/clinical pregnancy§	95% CI	0.85-1.08	0.86-1.14	0.50-8.83	0.19-3.90	0.14-25.35
	n	83	73		73	
	P	0.48	0.88		0.68	

* Variables adjusted for were, for 'number of good quality embryos transferred': type of hormonal ovulation stimulation, male infertility factor and man's age, and for the 'other IVF results': additionally type of infertility of the embryo recipient.

† Excluding 16 women without embryo transfer.

‡ Excluding 184 women without a positive pregnancy test result.

§ Excluding 191 women without a clinical pregnancy.

‖ P values in this column refer to the variation in IVF result explained by age as expressed in the likelihood ratio test, n refers to the sample size.

¶ Reference category: 30-34 years (OR = 1).

Table 4C. Odds ratios (OR) and 95% confidence intervals (CI) for crude and adjusted associations* between the woman's age to whom the embryos's are transferred (= $age_{recipient}$) and IVF outcomes for donor IVF

IVF outcome		C. Donor IVF and $age_{recipient}$ (N = 294)					
		Crude	Adjusted	In categories (years) Adjusted‖¶			
				24-34	35-39	45-49	50-59
Positive pregnancy	OR	1.01	1.01	0.95	0.59	1.01	0.85
test result	95% CI	0.97-1.04	0.96-1.06	0.29-3.10	0.23-1.47	0.48-2.13	0.31-2.30
/aspiration	n	293	252		252		
	P	0.72	0.69		0.80		
Clinical pregnancy	OR	1.02	1.01	1.04	0.59	1.20	0.94
/aspiration	95% CI	0.98-1.06	0.96-1.07	0.31-3.47	0.23-1.52	0.56-2.56	0.34-2.59
	n	293	252		252		
	P	0.37	0.63		0.70		
Ongoing pregnancy	OR	1.01	1.00	1.08	0.63	0.88	0.74
/aspiration	95% CI	0.97-1.05	0.94-1.06	0.29-4.03	0.22-1.78	0.38-2.00	0.24-2.21
	n	293	252		252		
	P	0.69	0.92		0.88		
No implantation	OR	0.98	0.99	0.76	1.41	0.76	0.94
/embryo transfer†	95% CI	0.95-1.02	0.94-1.05	0.20-2.91	0.54-3.69	0.35-1.65	0.33-2.65
	n	273	236		236		
	P	0.44	0.70		0.75		
Pregnancy loss	OR	1.00	1.03	0.65	0.87	1.18	1.18
/positive pregnancy	95% CI	0.93-1.07	0.94-1.13	0.06-6.52	0.15-5.11	0.31-4.51	0.19-7.45
test‡	n	96	85		85		
	P	0.97	0.53		0.97		
Spontaneous abortion	OR	1.02	1.05	0.63	0.81	1.68	1.75
/clinical pregnancy§	95% CI	0.94-1.10	0.95-1.16	0.05-8.35	0.11-6.18	0.42-6.75	0.25-12.25
	n	92	81		81		
	P	0.65	0.31		0.85		

* Variables adjusted for were, for 'number of good quality embryos transferred': type of hormonal ovulation stimulation, male infertility factor and man's age, and for the 'other IVF results': additionally type of infertility of the embryo recipient.

† Excluding 16 women without embryo transfer.

‡ Excluding 184 women without a positive pregnancy test result.

§ Excluding 191 women without a clinical pregnancy.

‖ P values in this column refer to the variation in IVF result explained by age as expressed in the likelihood ratio test, n refers to the sample size.

¶ Reference category: 40-44 years (OR = 1).

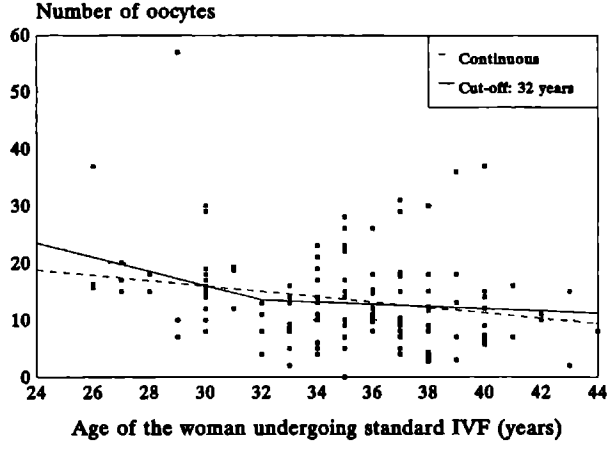


Figure 2. Number of oocytes per age of women who underwent standard IVF because of tubal pathology and who received a long GnRH-agonist ovulation stimulation protocol, and regression lines with and without cut-off point

Phase I: $Age_{provider}$ and IVF outcomes before embryo transfer

Both in standard IVF and in donor IVF the number of oocytes decreased steadily with increasing age of the woman who provided the oocytes (Table 3, columns A and B). For standard IVF, this association was best described ($F_{2,239} = 11.02$, $p = 0.0001$) by a model that allowed for a rapid decrease before the age of 32 years [ß = -1.25, SE(ß) = 0.38], followed by a slower decrease thereafter [ß = -0.19, SE(ß) = 0.17]. In Figure 2 data are shown from the women who underwent standard IVF because of tubal pathology, who received a long GnRH-agonist ovarian stimulation protocol. Both the accompanying regression lines are presented in the Figure, one using no cut-off point, and the better-fitting one which used a cut-off point at 32 years. For donor IVF, the number of oocytes donated by women aged 20 to 37 years decreased steadily by 0.24 oocyte per year.

The number of fertilized oocytes seemed to be negatively related to the age of the woman who provided the oocytes. However, this association disappeared after adjustment for confounding (Table 3, columns A and B).

In standard IVF, the transfer of at least one good quality embryo was not related to the age of the woman who provided the oocytes (Table 4A). In donor IVF a change in age effect best explained the variation in this IVF outcome [$-2\ln(L_1/L_2) = 11.49$, df = 2, p = 0.003]. At least one good embryo was transferred in 78% and 52% of the women who received oocytes from donors of 21 and 22 years, respectively (OR = 0.32, 95% CI 0.15-0.70). This occurred in only about 27% of those with $age_{provider}$ of 23 years and older (OR

= 1.01, 95% CI 0.92-1.10). All the women who underwent standard IVF were older than 23 years.

Phase II: $Age_{recipient}$ and IVF outcomes after embryo transfer

Next, analyses were performed to study whether the age of the women who received the embryos from donor IVF was related to pregnancy outcome. No association was found between $age_{recipient}$ and the probability of a positive pregnancy test result, clinical pregnancy, ongoing pregnancy, implantation, pregnancy loss and spontaneous abortion (Table 4C).

Phase III: $Age_{provider}$ and IVF outcomes after embryo transfer

As the aging effect of the recipient seemed to be negligible, the question remained as to whether the age of the oocyte provider affected these six pregnancy outcome parameters. On the basis of these data on donor IVF, it was assumed that there would be no effect of an aging uterus either in women who underwent standard IVF. This provided the opportunity to study the effect of $age_{provider}$ using data from both standard and donor IVF. The results of the analyses on $age_{provider}$ are shown in Tables 4A-B.

In standard and donor IVF, a positive pregnancy test result was not related to the age of the woman who provided the oocytes. After adjusting for confounding effects and after allowing for a change in trend, the probability of clinical pregnancy, as well as the probability of ongoing pregnancy after standard IVF only decreased steadily when the age of the oocyte provider exceeded 40 years. [For clinical pregnancy until age 40 years: OR = 1.01, 95% CI 0.92-1.12; after age 40 years: OR = 0.30, 95% CI 0.10-0.87; $-2\ln(L_1/L_2)$ = 10.37, df = 2, p = 0.006] [For ongoing pregnancy until age 40 years: OR = 1.01, 95% CI 0.91-1.12; after age 40 years: OR = 0.32, 95% CI 0.11-0.91; $-2\ln(L_1/L_2)$ = 9.00, df = 2, p = 0.01]. The probabilities of clinical pregnancy and ongoing pregnancy were 25-29% and 20-23% between the ages of 24 to 40, respectively. These probabilities decreased rapidly after age 40 years to 11% and 9%, respectively, and were even lower at higher ages. In donor IVF the $age_{provider}$ did not affect the probability of achieving a clinical or an ongoing pregnancy, but the women who donated the oocytes were 37 years or younger.

The probability of no implantation after embryo transfer in standard IVF increased with increasing woman's age (Table 4A). With higher age of the oocyte provider more often no implantation occurred. The variation in implantation was best explained by a logistic regression model that allowed for a change in trend after age 40 years [after adjustment for confounding effects: until age 40 years: OR = 1.03, 95% CI 0.93-1.14; after age 40 years: OR = 3.47, 95% CI 1.19-10.14; $-2\ln(L_1/L_2)$ = 12.32, df = 2, p = 0.002]. This resulted in a probability of no implantation that increased slightly from 70% at age 24 years to 79% at age 40. This probability increased drastically after age 40 years

to 93% at age 41 and higher at older age. Analogues to the absence of an $age_{provider}$ effect on the probability of achieving a clinical or an ongoing pregnancy in donor IVF, no age effect was found on implantation in women who underwent donor IVF (Table 4B), because all the women who provided oocytes were younger than 40 years.

In women who underwent standard or donor IVF, no $age_{provider}$ trend was apparent for the occurrence of pregnancy loss after a positive pregnancy test, or for the probability of spontaneous abortion after a clinical pregnancy (Tables 4A-B).

Discussion

Several studies have addressed the effect of woman's age on the success of IVF treatment with donated oocytes (e.g. Sauer *et al*,[12,16] Abdalla *et al*,[13] Flamigni *et al*;[19] Navot *et al*;[14] Check *et al*;[15] Cano *et al*;[21] Borini *et al*.[20]) The results were contradictory. The influence of confounding factors, the use of arbitrary cut-off points, small numbers of women under study and a narrow age range might have caused the disagreement. Therefore we performed this study. We used an extended data base of the donor IVF cycles described by Sauer *et al*[12,16] and Legro *et al*.[17] In contrast with these previous reports and reports by others, we studied the age effect continuously, adjusted for confounding effects and used only the first IVF treatment cycle to prevent bias from repeated cycles after a failed treatment cycle.

Although the data base used in this study contained a relatively large number of couples who underwent standard or donor IVF, the number might still be too small to demonstrate statistically significant age effects. Therefore, we did not interpret the p-values too rigidly and we attached more importance to the values and the directions of the regression coefficients.

The importance of adjusting for confounding is illustrated by the results found for fertilized oocytes. The number of fertilized oocytes was related to the $age_{provider}$ only when no adjustment was made for the effects of the stimulation protocol, infertility factor of the women who underwent standard IVF, type of IVF, male infertility factor and male age. Thus, the relationship between the number of fertilized oocytes and $age_{provider}$ could be explained by the unequal distribution of the confounding factors over $age_{provider}$. Although we presume that we were able to adjust for confounding effects of the most important confounders, there might have been residual confounding. Examples of such factors might have been the number of ovaries and the basal follicle stimulating hormone (FSH) levels of the oocyte providers. Unfortunately, this information was not available.

We categorized the women according to age only to see whether there was a change in age effect. Where appropriate, we used the age distribution on a continuous scale to

describe the age effect and allowed for a change in age effect by incorporating a cut-off point. Some changes in the age effect were found at a certain age in the women who underwent standard IVF. The number of oocytes retrieved declined rapidly by 1.25 oocyte per year between 24 and 32 years of age and by only 0.19 oocyte per year after the age of 32 years. In the women who underwent standard IVF, the probability of achieving a clinical or an ongoing pregnancy per aspiration decreased and the probability of no implantation after embryo transfer increased sharply after the age of 40 years. Although the model provided the best explanation for the variation in IVF results at the specific cut-off points, models that used cut-off points of one year higher or lower were also reasonably accurate. Therefore, these cut-off points should not be considered too strictly. The authors of a recent study on standard IVF reported that they could not demonstrate a change in age effect.[25] However, as they included age only as a (log)linear variable in the model, it was not possible (by definition) to find a change in trend. In other studies, changes in the age effect were reported. A decline of about 12% per year in the probability of conception after treatment with artificial insemination with donor semen only occurred after the age of 31 years.[5] This cannot be compared directly to our age effect results, because of the difference in treatment, especially the ovarian hyperstimulation. For example, it is possible that controlled ovarian hyperstimulation delays the appearance of this age effect. Two studies on standard IVF found a gradual decrease in positive pregnancy test results or ultrasound evidence of a gestational sac and live birth rates after the age of 34 years[9] and a steeper decline in the pregnancy rate after the age of 37 than before the age of 37 years.[7]

Although the age of the oocyte provider was negatively related to the number of oocytes retrieved, there was no effect on the occurrence of a positive pregnancy test. A decrease in the ultimate aim of IVF treatment, to achieve an ongoing pregnancy, only occurred after the oocyte provider had reached 40 years of age. This was also reflected by the lack of implantation after embryo transfer and no clinical pregnancy. These effects were only found in the women who underwent standard IVF, but not in donor IVF. This might be explained by the age distribution of the oocyte donors: they were all younger than 38 years. Such findings might indicate an aging affect of the oocytes, but we consider the possibility that the effects were due to a decrease in the number of oocytes with age, without any loss of oocyte quality in older women. To determine whether this hypothesis holds, we studied in couples without male infertility factor who underwent standard IVF, whether the occurrence of an ongoing pregnancy was still related to the $age_{provider}$ after controlling for the number of oocytes retrieved in addition to adjusting for the confounding factors. There was still a decrease in the probability of achieving an ongoing

pregnancy after the age of 40 years [until age 40 years: OR = 1.02, 95% CI 0.90-1.15; after age 40 years: OR = 0.31, 95% CI 0.11-0.89; $-2\ln(L_1/L_2) = 9.03$, df = 2, p = 0.01]. Moreover, the amount of variation in the occurrence of an ongoing pregnancy that could be explained by $age_{provider}$ was similar in the model which included the number of oocytes and in the model which excluded the number of oocytes in couples without a male infertility factor [$-2\ln(L_1/L_2) = 8.85$, df = 2, p = 0.01]. Thus, the effect of $age_{provider}$ was not caused by the number of oocytes retrieved.

A larger proportion of the women who received oocytes from donors aged 20-23 years had at least one good quality embryo transferred than those who received oocytes from older women. This age effect was not found in standard IVF and might be explained by the age distribution of the women who underwent standard IVF: all the women were 24 years or older. This seems to indicate that the younger the oocyte donor, the more oocytes that will be retrieved and the higher the probability that at least one good quality embryo will be transferred. When we examined this in couples without a male infertility factor who underwent donor IVF, the observed relationship still remained after controlling for the number of oocytes [until age 23 years: OR = 0.28, 95% CI 0.11-0.68, after age 23 years: OR = 1.01, 95% CI 0.90-1.13; $-2\ln(L_1/L_2) = 12.78$, df = 2, p = 0.002]. The quality of the oocytes from young oocyte donors seemed to be better than that from older donors.

This study shows that there is no aging effect of the uterus, but that there is an aging effect of the oocyte in women of older than 40 years.

References

1. Döring GK. The incidence of anovulatory cycles in women. J Reprod Fert 1969;Suppl. 6:77-81.
2. Metcalf MG, Donald RA, Livesey JH. Classification of menstrual cycles in pre- and perimenopausal women. J Endocrinol 1981;91:1-10.
3. Schwartz D, Mayaux MJ. Female fecundity as a function of age. Results of artificial insemination in 2193 nulliparous women with azoospermic husbands. Fédération CECOS. N Engl J Med 1982;306:404-406.
4. Virro MR, Shewchuk AB. Pregnancy outcome in 242 conceptions after artificial insemination with donor sperm and effects of maternal age on the prognosis for successful pregnancy. Am J Obstet Gynecol 1984;148:518-524.
5. Van Noord-Zaadstra BM, Looman CWN, Alsbach H, Habbema JDF, Te Velde ER, Karbaat J. Delaying childbearing: effect of age on fecundity and outcome of pregnancy. Br Med J 1991;302:1361-1365.
6. Padilla SL, Garcia JE. Effect of maternal age and number of in vitro fertilization procedures on pregnancy outcome. Fertil Steril 1989;52:270-273.
7. Piette C, De Mouzon J, Bachelot A, Spira A. In-vitro fertilization: influence of women's age on pregnancy rates. Hum Reprod 1990;5:56-59.
8. Chetkowski RJ, Rode RA, Burruel V, Nass TE. The effect of pituitary suppression and the women's age on embryo viability and uterine receptivity. Fertil Steril 1991;56:1095-1103.

9. Tan SL, Royston P, Campbell S, Jacobs HS, Betts J, Mason B, Edwards RG. Cumulative conception and livebirth rates after in-vitro fertilisation. Lancet 1992;339:1390-1394.
10. Hull MGR, Fleming CF, Hughes AO, McDermott A. The age-related decline in female fecundity: a quantitative controlled study of implanting capacity and survival of individual embryos after in vitro fertilization. Fertil Steril 1996;65:783-790.
11. Abdelmassih R, Sollia S, Moretto M, Acosta AA. Female age is an important parameter to predict treatment outcome in intracytoplasmic sperm injection. Fertil Steril 1996;65:573-577.
12. Sauer MV, Paulson RJ, Lobo RA. Reversing the natural decline in human fertility. An extended clinical trial of oocyte donation to women of advanced reproductive age. JAMA 1992;268:1275-1279.
13. Abdalla HI, Burton G, Kirkland A, Johnson MR, Leonard T, Brooks AA, Studd JWW. Age, pregnancy and miscarriage: uterine versus ovarian factors. Hum Reprod 1993;8:1512-1517.
14. Navot D, Drews MR, Bergh PA, Guzman I, Karstaedt A, Scott RT, Garrisi GJ, Hofmann GE. Age-related decline in female fertility is not due to diminished capacity of the uterus to sustain embryo implantation. Fertil Steril 1994;61:97-101.
15. Check JH, Askari HA, Fisher C, Vanaman L. The use of a shared donor oocyte program to evaluate the effect of uterine senescence. Fertil Steril 1994;61:252-256.
16. Sauer MV, Paulson RJ, Ary BA, Lobo RA. Three hundred cycles of oocyte donation at the University of Southern California: assessing the effect of age and infertility diagnosis on pregnancy and implantation rates. J Assist Reprod Genet 1994;11:92-96.
17. Legro RS, Wong IL, Paulson RJ, Lobo RA, Sauer MV. Recipient's age does not adversely affect pregnancy outcome after oocyte donation. Am J Obstet Gynecol 1995;172:96-100.
18. Levran D, Ben-Shlomo I, Dor J, Ben-Rafael Z, Nebel L, Mashiach S. Aging of endometrium and oocytes: observation on conception and abortion rates in an egg donation model. Fertil Steril 1991;56:1091-1094.
19. Flamigni C, Borini A, Violini F, Bianchi L, Serrao L. Oocyte donation: comparison between recipients from different age groups. Hum Reprod 1993;8:2088-2092.
20. Borini A, Bianchi L, Violini F, Maccolini A, Cattoli M, Flamigni C. Oocyte donation program: pregnancy and implication rates in women of different ages sharing oocytes from single donor. Fertil Steril 1996;65:94-97.
21. Cano F, Simón C, Remohí J, Pellicer A. Effect of aging on the female reproductive system: evidence for a role of uterine senescence in the decline in female fecundity. Fertil Steril 1995;64:584-589.
22. Meldrum DR. Female reproductive aging - ovarian and uterine factors. Fertil Steril 1993;59:1-5.
23. Sauer MV, Ary BR, Paulson RJ. The demographic characteristics of women participating in oocyte donation: a review of 300 consecutively performed cycles. Int J Gynaecol Obstet 1994;45:147-151.
24. Paulson RJ, Sauer MV, Lobo RA. Embryo implantation after human in vitro fertilization: importance of endometrial receptivity. Fertil Steril 1990;53:870-874.
25. Roseboom TJ, Vermeiden JPW, Schoute E, Lens JW, Schats R. The probability of pregnancy after embryo transfer is affected by age of the patient, cause of infertility, number of embryos transferred and the average morphology score, as revealed by multiple logistic regression analysis. Hum Reprod 1995;10:3035-3041.

4. Discussion

The studies described in this thesis were performed to gain greater insight into human fecundity under natural conditions and in the case of treatment with in vitro fertilization (IVF). Interest in this subject was nourished by the hypothesis that in situations of diminished fecundity, the probability of poor reproductive outcome is increased.[1] Various topics are presumed to cause variation in fecundity. In this thesis, the main focus is on the season of the year and the characteristics of patients undergoing IVF treatment, such as the woman's age and indications for treatment.

In this chapter, studies on seasonality in fecundity are discussed to reach a conclusion about the actual effect of the season on fecundity. Subsequently, the consequences are considered for the studies on seasonality in poor reproductive outcome. Next, the influence of the woman's age on fecundity is evaluated under natural circumstances and during IVF treatment. Prognostic models are used to address the study on variation in fecundity during IVF treatment. The importance of prognostic models and their limits are discussed. Attention is paid to the quality of reproductive outcome after in vitro fertilization. The impact of this thesis on clinical management and further research is considered.

Seasonality in fecundity and poor reproductive outcome

Whether the season influences fecundity is addressed in this thesis by studying ovulation, the time to pregnancy and IVF outcomes. No seasonal pattern was found in ovulations in our study (Chapter 2.2). Two other studies observed seasonality in ovulations in women living under extreme situations of photoperiodicity[2] or temperature.[3] It might be concluded that the seasonal influence on a female's potential to conceive is negligible in countries such as the Netherlands with less extreme variation in the photoperiod and temperature. A bimodal seasonal pattern was found in the time to pregnancy in French-Canadian women living in the 17th and 18th centuries, indicating the most fecund periods during December-January and June-July. The difference in waiting time was small (Chapter 2.4.2).

We also studied whether there were seasonal patterns in IVF results by using data from the University Hospital Nijmegen. Seasonality was apparent in consecutive IVF outcomes: in the number of oocytes retrieved; additionally, in the patients from whom at least one oocyte was retrieved seasonality was also apparent in the fertilization rate, embryo quality, pregnancy rate and birth rate. The best results were seen in November-February

(Chapter 2.3). Conceivably, the quality of the oocyte or sperm was better during this period. Another research group published their results at the same time on a study performed using data from couples in whom the male partner had sufficient sperm function. They found the least fertilization in IVF cycles during July-September and considered this to be a coincidence as it was strictly caused by a low fertilization rate in July, while no such pattern was reflected in other IVF outcomes.[4] Another study found no seasonal variation in pregnancy rate after the transfer of cryopreserved embryos during an unstimulated cycle.[5] In our prognostic models for predicting an ongoing pregnancy after IVF treatment, the season did not fulfil the criteria for inclusion into the models, i.e. a p value of <0.10 and a change in c index of >0.005 (Chapter 3.2). These three latter studies indicate that another reason for the seasonal pattern found in our former study (Chapter 2.3) is more likely, namely changes in IVF treatment protocol during the year. For developing the prognostic models, only the IVF treatments that had been carried out since March 1991 were analysed, because no major changes had been introduced in the treatment protocol from that time on. However, treatments before March 1991 were included in the study on seasonality in IVF outcome.

Thus, from these studies on seasonality in fecundity it might be concluded that seasonal variation in human fecundity is negligible nowadays in countries such as the Netherlands with moderate photoperiodicity, moderate temperature changes and minor seasonal variation in energy intake and energy expenditure. The consequence of this conclusion for studies on seasonality in poor reproductive outcome is that if a seasonal pattern in poor reproductive outcome is found in countries similar to the Netherlands, it is unlikely to be the result of variation in fecundity. With this in mind, it is not surprising that a critical review of studies on seasonality in Down syndrome at birth showed very inconsistent patterns (Chapter 2.5).

Influence of the woman's age on fecundity under natural circumstances and during IVF treatment

Under natural circumstances, it is well-known that a female's fecundity decreases with increasing age. This phenomenon was also reflected in the time to pregnancy (Chapters 2.4.2 and 2.4.3). Moreover, fecundity was expected to be reduced during the first years after the menarche. Our large-scale study on seasonality in the interval between marriage and the first birth in French-Canadian women in the 17th and 18th centuries supported this hypothesis (Chapter 2.4.3). The studies on the time to pregnancy in Denmark (Chapter 2.4.2), however, did not confirm these findings, probably because the number of young women was too small. In the case of IVF, the aging effect on pregnancy

was the most obvious in women after the age of about 40 years who used their own oocytes. As the reduction in the pregnancy rate after IVF could not be explained by a lower number of oocytes retrieved after controlled hormonal ovulation induction, the age effect seemed to be due to a decrease in oocyte quality (Chapter 3.4). This result gives additional reason to reconsider the age criterion for IVF treatment, as cost-efficiency will decrease after the age of forty. Moreover, decreased quality of the oocyte in these women might affect the quality of the offspring.

Variation in fecundity during IVF treatment

IVF treatment is obviously meant to increase fecundity, but the probability of achieving a livebirth after IVF treatment varies greatly between patients. Prognostic models were used to determine which characteristics of patients and treatment modalities can predict fecundity during IVF treatment. Such prognostic studies are of great practical importance, because the outcome of an IVF treatment is uncertain and it makes high physical and psychological demands, involves high financial cost and can have serious side-effects. The results of prognostic studies can provide patients and physicians with information about the likelihood of pregnancy, which will facilitate decisions. Moreover, these models might help policy makers to specify the indications for IVF treatment.

Before prognostic models can be implemented in clinical practice, their external validity must be demonstrated. The external validity of untested or only internally validated prognostic models must be seriously questioned (Chapter 3.2). In Chapter 3.2, a description is given of the development of prognostic models with data from the University Hospital Nijmegen and their application to data from the Catharina Hospital Eindhoven. The overall predictive performance of the models was low, except for a model that predicted ongoing pregnancy at the moment of embryo transfer. Although this model has no clinical value, it does provide clues about an explanation for the poor results of the other models: differences in treatment between the two centres might have decreased the predictive accuracy. This formed one of the topics addressed in a subsequent study (Chapter 3.3). In patients from Eindhoven who did not have the first IVF treatment cancelled because of too many follicles, couples could be identified that had only a 0-10% probability of achieving an ongoing pregnancy in the second IVF treatment. A further test of the models with data from the Diaconessenhuis Voorburg showed poor discriminative performance. Thus the external validity of the models is still questionable. To test whether the models performed well at the clinic where they were developed, new data from the University Hospital Nijmegen were used. The results were inconclusive, because of the small number of women with a low predicted probability of achieving an ongoing

pregnancy. Thus more data and further testing are required.

Poor reproductive outcome after IVF treatment

The hypothesis that IVF treatment increases the risk of poor reproductive outcome prompted several follow-up studies. Apart from an increased risk of poor reproductive outcome because of a high rate of multiple births, several of these studies reported that there was no increased risk of poor reproductive outcome after a standard IVF procedure [e.g.[6-8]]. There are, however, reports on singleton births after IVF that gave reason for concern, because of a relatively high prevalence of preterm births[9-15] and (very) low birth weight for gestational age.[11-14,16,17] It remains to be seen whether these differences were still be present after adjusting for the presumably higher frequency of unfavourable characteristics of the women undergoing IVF.[12,17] One study reported increased perinatal mortality,[9] but other large studies could not confirm this finding.[11,14] Increased perinatal mortality rates were found in spina bifida and transposition of the great vessels,[18] minor congenital malformations[15] and major congenital malformations;[19] but in large studies, no overall increased risk of minor or major congenital malformations was found.[9-11] Children with a very low birth weight, who are small-for-gestational age or are born prematurely, are at increased risk for developmental problems.[20-22] As these situations occur more frequently after IVF than after normal conceptions, a high prevalence of developmental problems can be expected in both singleton and multiple births after IVF. However, studies have shown that the mental and physical development of children born after IVF were not delayed in comparison with normally-conceived children, when amongst others, birth weight, gestational age and multiple deliveries were taken into account;[23] this remained valid regardless of birth weight and gestational age.[24] Other authors showed that IVF children had even higher scores on developmental scales than the standards.[25,26] The frequency of poor reproductive outcome after cryopreservation of the embryos was not increased in comparison with fresh embryo transfer.[27] But again, there was a higher frequency of low birth weight and preterm births in these children compared to normally-conceived children.[28] The conclusion of Sutcliffe *et al*[29] that the proportion of minor congenital malformations was similar in children born after IVF with cryopreserved embryo transfer and in children conceived normally was unconvincing: minor congenital malformations occurred in 32.4% and 22.2% of the children, respectively (odds ratio = 1.7, 95% confidence interval 0.8-3.5). The quality of the reproductive outcome after intracytoplasmatic sperm injection (ICSI) also seems to be similar to that after standard IVF.[30] However, the potential genetic risk in men who are offered ICSI still forms a reason for concern and the treatment itself may carry further risks.[31-33] It is necessary to

implement this technique conservatively, perform genetic screening of male patients and to follow-up subsequent offspring to study whether developmental and physical problems exist. Problems are especially likely to arise when ICSI is used in combination with spermatozoa retrieved from the testicles (TESE) or epididymis (MESA), because increased genetic aberrations may occur in sperm as a consequence of a genetic cause of subfertility and because immature gametes may be retrieved in the case of TESE or postmature gametes in the case of MESA.

Consequences for clinical management

It is concluded that there is no or negligible seasonal variation in human fecundity nowadays in countries such as the Netherlands with moderate photoperiodicity and temperature changes and with minor seasonal variation in energy intake and energy expenditure. The season is no longer an important factor, if it ever was, for influencing human fecundity in countries such as these.

Valid predictions for an ongoing pregnancy after IVF are not yet possible, although there may be a few exceptions. In general, IVF physicians should not use the models described in Chapter 3.2 and Chapter 3.3, or inadequately validated models from the literature. Until prognostic models have made reliable predictions in other populations, the risk that they will give false predictions is too large.

Consequences for further research

Nowadays, seasonal variation in human fecundity is absent or negligible in countries such as the Netherlands with only moderate photoperiodicity, temperature changes and seasonal variation in energy intake and energy expenditure. Therefore, seasonality in poor reproductive outcome, because of seasonality in fecundity, is not to be expected. Research into this topic should focus on populations living under more extreme conditions. Even if seasonality in fecundity does become apparent, it will not be possible to formulate a specific hypothesis about low- and high-risk months for poor reproductive outcome, because such a hypothesis would have to take into account the seasonal pattern in ovulation, in sperm characteristics, perhaps in fertilization and nidation, early pregnancy loss and spontaneous abortions.

In the Introduction, other natural situations of subfecundity are mentioned: just after discontinuation of the use of oral contraceptives, after pregnancy, during and after lactation, with an endocrine disease, in the case of extreme under- or overweight and during weight changes. Whether these and other natural conditions of subfecundity cause poor reproductive outcome remains to be studied, but valid epidemiological studies

concerning these topics are very difficult to realize (Chapter 1).

The development of valid prognostic models for IVF outcome is hampered by the lack of consensus in treatment and registration and by the small numbers of patients available for study. For instance, there are no uniform indications for IVF treatment and the controlled hormonal ovulation induction protocols differ between centres. The characteristics of the patients, the performed treatment techniques and the outcomes should be registered and entered into a computer prospectively to achieve a database of high quality. All IVF treatments, whether they resulted in oocyte aspiration or not, must be included. The reasons for the discontinuation of treatment should be registered as well, to enable the estimation of realistic cumulative pregnancy rates (Chapter 3.1). Compiling a data registry in retrospect by using information form the patients' charts is vulnerable to incompleteness and errors. Consensus meetings for IVF treatment and prospective data gathering on a national, or even on an international basis, are recommended. The availability of high quality databases will enable the further development of prognostic models to predict, for instance, the cumulative probability of achieving a livebirth after successive IVF cycles and the probability that IVF treatment will result in poor reproductive outcome.

Former studies have indicated increased prevalences of poor reproductive outcome, such as small-for-gestational age and preterm birth, in children born after IVF, which were only partly due to an increase in multiple pregnancies. Several follow-up studies showed no impaired mental or physical development in children born after standard IVF. Further studies on developmental problems in children should focus on other IVF modalities, such as the transfer of cryopreserved embryos and fertilization using ICSI, especially in cases of severe male subfertility that require spermatozoa collection from the testicles or epididymis. Furthermore, it is highly recommended that each new procedure implemented as part of the IVF treatment should be tested before it is widely used on subfecund populations. Such tests should focus on both the efficiency of the treatment and the development of children born after the infertility treatment.

References

1. Jongbloet PH. The effect of preovulatory overripeness of human eggs on development. In: Blandau RJ, (Ed.). Aging Gametes. Basel: S. Karger AG, The effect of preovulatory overripeness of human eggs on development. 1975;300-329.

2. Timonen S, Franzas B, Wichmann K. Photosensibility of the human pituitary. Ann Chir Gynaec Fenn 1964;53:165-172.

3. Rameshkumar K, Thomas JA, Mohammed A. Atmospheric temperature & anovulation in south Indian women with primary infertility. Indian J Med Res 1992;96:27-28.

4. Fleming C, Nice L, Hughes AO, Hull MGR. Apparent lack of seasonal variation in implantation

rates after in-vitro fertilization. Hum Reprod 1994;9:2164-2166.
5. Dunphey BC, Anderson-Sykes S, Brant R, Pattinson HA, Greene CA. Human embryo implantation following in-vitro fertilization: is there a seasonal variation? Hum Reprod 1995;10:1825-1827.
6. Shoham Z, Zosmer A, Insler V. Early miscarriage and fetal malformations after induction of ovulation (by clomiphene citrate and/or human menotropins), in vitro fertilization, and gamete intrafallopian transfer. Fertil Steril 1991;55:1-11.
7. Rizk B, Doyle P, Tan SL, *et al.* Perinatal outcome and congenital malformations in in-vitro fertilization babies from the Bourn-Hallam group. Hum Reprod 1991;6:1259-1264.
8. Alsalili M, Yuzpe A, Tummon I, *et al.* Cumulative pregnancy rates and pregnancy outcome after in-vitro fertilization: >5000 cycles at one center. Hum Reprod 1995;10:470-474.
9. Saunders DM, Lancaster P. The wider perinatal significance of the Australian in vitro fertilization data collection program. Am J Perinatol 1989;6:252-255.
10. MRC working party on children conceived by in vitro fertilisation. Births in Great Britain resulting from assisted conception, 1978-87. Br Med J 1990;300:1229-1233.
11. Tan SL, Doyle P, Campbell S, *et al.* Obstetric outcome of in vitro fertilization pregnancies compared with normally conceived pregnancies. Am J Obstet Gynecol 1992;167:778-784.
12. Olivennes F, Rufat P, André B, Pourade A, Quiros MC, Frydman R. The increased risk of complication observed in singleton pregnancies resulting from in-vitro fertilization (IVF) does not seem to be related to the IVF method itself. Hum Reprod 1993;8:1297-1300.
13. Wang JX, Clark AM, Kirby CA, *et al.* The obstetric outcome of singleton pregnancies following in-vitro fertilization/gamete intra-fallopian transfer. Hum Reprod 1994;9:141-146.
14. FIVNAT. Pregnancies and births resulting from in vitro fertilization: French national registry, analysis of data 1986 to 1990. Fertil Steril 1995;64:746-756.
15. Verlaenen H, Cammu H, Derde MP, Amy JJ. Singleton pregnancy after in vitro fertilization: expectations and outcome. Obstet Gynecol 1995;86:906-910.
16. Doyle P, Beral V, Maconochie N. Preterm delivery, low birth weight and small-for-gestational age in liveborn singleton babies resulting from in-vitro fertilization. Hum Reprod 1992;7:425-428.
17. Petersen K, Hornnes PJ, Ellingsen S, *et al.* Perinatal outcome after in vitro fertilisation. Acta Obstet Gynecol Scand 1995;74:129-131.
18. Lancaster PAL. Congenital malformations after in-vitro fertilisation (letter). Lancet 1987;ii:1392-1393.
19. Licata D, Garzena E, Mostert M, Farinasso D, Fabris C. Congenital malformations in babies born after assisted conception (letter). Paediatr Perinat Epidemiol 1993;7:222-223.
20. Veen S, Ens-Dokkum MH, Schreuder AM, Verloove-Vanhorick SP, Brand R, Ruys JH. Impairments, disabilities, and handicaps of very preterm and very low-birthweight infants at five years of age. From the collaborative project on preterm and small for gestational age infants (POPS) in the Netherlands. Lancet 1991;338:33-36.
21. Halsey AL, Collin MF, Anderson CL. Extremely low birth weight children and their peers: a comparison of preschool performance. Pediatrics 1993;91:807-811.
22. Roussounis SH, Hubley PA, Dear PRF. Five-year-follow-up of very low birthweight infants: neurological and psychological outcome. Child Care Health Dev 1993;19:45-59.
23. Brandes JM, Scher A, Itzkovits J, Thaler I, Sarid M, Gershoni Baruch R. Growth and development of children conceived by in vitro fertilization. Pediatrics 1992;90:424-429.
24. Morin NC, Frank LM, Chee EM. Congenital malformations and psychosocial development in children conceived by in vitro fertilization. J Pediatr 1989;115:222-227.
25. Yovich JL, Parry TS, French NP, Grauaug AA. Developmental assessment of twenty in vitro

fertilization (IVF) infants at their first birthday. J In Vitro Fert Embryo Transf 1986;3:253-257.
26. Mushin DN, Barreda-Hanson MC, Spensley JC. In vitro fertilization children: early psychosocial development. J In Vitro Fert Embryo Transf 1986;3:247-252.
27. Wada I, Macnamee MC, Wick K, Bradfield JM, Brinsden PR. Birth characteristics and perinatal outcome of babies conceived from cryopreserved embryos. Hum Reprod 1994;9:543-546.
28. Sutcliffe AG, D'Souza SW, Cadman J, Richards B, McKinlay IA, Lieberman B. Outcome in children from cryopreserved embryos. Arch Dis Childh 1995;72:290-293.
29. Sutcliffe AG, D'Souza SW, Cadman J, Richards B, McKinlay IA, Lieberman B. Minor congenital anomalies, major congential malformations and development in children conceived from cryopreserved embryos. Hum Reprod 1995;10:3332-3337.
30. Wisanto A, Magnus M, Bonduelle M, *et al.* Obstetric outcome of 424 pregnancies after intracytoplasmatic sperm injection. Hum Reprod 1995;10:2713-2718.
31. Patrizio P. Intracytoplasmatic sperm injection (ISCI): Potential genetic concerns. Hum Reprod 1995;10:2520-2523.
32. Hollanders JMG, Meuleman EJH, Wetzels AMM. Risico's van intracytoplasmatische zaadcelinjectie voor het nageslacht? Ned Tijdschr Geneeskd 1996;140:9-11.
33. Baschat AA, Küpker W, Al Hasani S, Diedrich K, Schwinger E. Results of cytogenetic analysis in men with severe subfertility prior to intracytoplasmatic sperm injection. Hum Reprod 1996;11:330-333.

Summary

This thesis addresses human fecundity, the ability to conceive within one menstrual cycle when pregnancy is desired and no contraceptive methods are used. A combination of two approaches has been used, namely studying natural conditions and situations during infertility treatment. The core questions were: 1) Is there seasonal variation in human fecundity? and 2) Which combinations of factors have prognostic value regarding the results of treatment with in vitro fertilization (IVF)? Such knowledge might give directions to prevention programmes for subfertility and the allocation of infertility treatments.

Seasonality in human fecundity was expected because there is seasonal variation in reproduction in most mammals and several studies on humans have indicated seasonal patterns in hormone levels associated with reproduction, ovulation, sperm characteristics, the probability of becoming pregnant, the time to pregnancy and births. These studies did not find a consistent seasonal pattern, however. There might be a trend towards higher fecundity in women during the 'light season', whereas the fertility of men seems to be increased during the 'dark season'.

The first studies in this thesis concern methodological issues related to epidemiological studies on seasonality in reproductive outcome. One problem in comparing results of such studies is the variety in parameters used to describe this. Three frequently used parameters are the prevalence, the index and the ratio observed versus expected. The validity and precision of these three parameters are the same, only their levels of convenience differ. The statistical feasibility of the prevalence is the best, therefore this parameter has been used throughout the thesis (Chapter 2.1.1). The application of a regression analysis using a cosine function is described for testing seasonal patterns, which allows for adjustment of confounding effects (Chapter 2.1.2). Whether seasonality in pregnancy planning can confound a relation between a seasonally distributed exposure and poor reproductive outcome was studied by means of simulations. Under realistic circumstances this confounding was negligible (Chapter 2.1.3).

Seasonality in ovulatory cycles was studied using data from 407 women with menstrual cycles that were shorter than 6 weeks who had visited the fertility clinic at the University Hospital Nijmegen for the first time in 1991 or 1992. To establish whether ovulation had occurred serial transvaginal ultrasound was performed and midluteal progesterone levels were measured during one menstrual cycle. No seasonal variation was

found. It is likely that changes in the photoperiod, environmental temperature and in seasonal energy intake and energy expenditure are too small in the Netherlands to cause any detectable seasonal variation in ovulation nowadays, as two other studies found a seasonal pattern in ovulation under more extreme circumstances (Chapter 2.2).

Some seasonal variation was found in the results of the first IVF cycles of 1126 couples treated at the University Hospital Nijmegen between 1987 and 1993. The best results were achieved when treatments were started in November-February (Chapter 2.3). However, when only the data were used from first IVF treatments between 1991 and 1994 (N = 757) to develop prognostic models for ongoing pregnancy, the season did not show any predictive value (Chapter 3.2). This indicates that seasonal changes in treatment protocols may have occurred in the earlier years of IVF treatment at this hospital.

To study seasonality in the time to pregnancy, data were used from three Danish studies in which pregnant women (N = 3,657) and textile workers or pharmacy assistants who had been pregnant (N = 1,053 and N = 734, respectively) were asked about the time it had taken for them to conceive. A higher probability of a prolonged time to pregnancy was found before conceptions in February-April and a lower probability before conceptions in August-October. However, this pattern might have been confounded by seasonality in pregnancy planning (Chapter 2.4.1). Therefore another study was performed using data from a population registry containing information about 18.970 French-Canadian women who married for the first time in the 17th or 18th century. The time to pregnancy was approximated by the interval between marriage and the first birth minus 38 weeks. This study showed a minor bimodal seasonal pattern in the time to pregnancy, with the most fecund periods during December-January and June-July. Moreover, it was illustrated that a large bias in seasonality in the time to pregnancy can occur if the season is defined by the month of conception and seasonal variation in pregnancy planning is apparent (Chapter 2.4.2).

About 70% of all Down syndrome are caused by maternal nondisjunction during the first meiosis, which occurs in the last few days before ovulation. Seasonal influence of fecundity might have an influence on this stage of oocyte ripening, because of intervention with hormone production by the hypothalamus-pituitary-ovarian axis. No consistent seasonal pattern was found in a critical review of twenty studies on seasonality in the prevalence of Down syndrome at birth which were published during the period 1966-1995. It is unknown whether a seasonal pattern in Down syndrome at the moment of conception disappears because of pregnancy loss. This could not be studied with the data available (Chapter 2.5).

IVF treatment is meant to increase fecundity. The cumulative pregnancy rates that have been presented in the literature often give a too optimistic impression of the fecundability achieved after successive IVF cycles. This overestimation will occur if one assumes that the couples who stop treatment without having achieved a pregnancy have the same probability of becoming pregnant as those who continue. A common reason for discontinuation of IVF treatment is, however, poor results. A more realistic estimation of the cumulative pregnancy rate is recommended, which takes into account the reasons for stopping treatment. This was illustrated by using the data from 872 couples who were treated for the first time with IVF at the University Hospital Nijmegen between 1988 and 1992. In this population, the estimated cumulative ongoing pregnancy rates after five initiated IVF cycles was 30% under the most conservative assumptions, 56% under the optimistic assumption which is often used in the literature and 34% when the reason for stopping treatment was taken into account (Chapter 3.1).

Prognostic modelling was used to asses which combinations of patient characteristics indicate low or high fecundability during IVF treatment. Information for patients about their probability of success with IVF treatment can be improved using prognostic models and they can assist physicians when counselling a patient. Prognostic models to predict the probability of achieving an ongoing pregnancy at several stages during IVF treatment were developed using data from the University Hospital Nijmegen 1991-1994 (N = 757). The models were applied to data from the Catharina Hospital Eindhoven (N = 432). The only prognostic factors at the start of the first IVF treatment regarding the probability of achieving an ongoing pregnancy during the first IVF treatment were the age of the woman and prior gestations. Prognostic factors for predicting ongoing pregnancy during the second treatment were: the age of the woman, idiopathic infertility and embryo transfer during the first IVF treatment. These two models which made predictions at the start and at the end of the first IVF cycle did not show any predictive value in the Eindhoven population. However, a model that made predictions at the time of embryo transfer performed fairly well. Although this model has no clinical meaning, it suggests that the poor predictive performance of the other two models may arise because of differences in the ovulation stimulation protocols and in the decisions for cancelling the treatment because of the risk of ovarian hyperstimulation (Chapter 3.2). In a subsequent study (Chapter 3.3), it was found that these models also made poor predictions of the probability of achieving an ongoing pregnancy in IVF patients at the University Hospital Nijmegen in 1995 and 1996 (N = 208) and at the Diaconessenhuis Voorburg (N = 1,424), even after calibration of the model by using a shrinkage estimator to adjust for overfitting of the models. However, for Nijmegen the testing was imprecise because of the small number

of patients with a low predicted probability. When the women in the Eindhoven population in whom the first IVF cycles were cancelled because of too many follicles were disregarded, it became possible to identify patients with a maximum probability of at the most 10% of achieving an ongoing pregnancy during the second IVF cycle. This applied to one third of the Eindhoven population. Thus as the prognostic models do not show good external validity, widespread use in clinical practice is not yet recommended, although there might be exceptions for specific situations. In addition, prognostic models from the literature that have been inadequately validated should not be used in clinical practice, as they will lead too often to false predictions.

It has been known for a long time that the fecundity of women decreases with increasing age. Using data from couples who underwent standard IVF treatment (N = 277) or IVF treatment with donor oocytes (N = 294) at the University of Southern California, Los Angeles, the USA, such a decrease in fecundity was also noticed during IVF treatment when the woman's own oocytes were used, but this could be solved by using donor oocytes. An effect of aging of the uterus did not seem to occur. In women who used their own oocytes there was a steep decrease in the probability of achieving an ongoing pregnancy after the age of forty years. As this lower probability could not be explained by the lower number of oocytes retrieved from these women, it indicates decreasing oocyte quality in older women (Chapter 3.4).

The following can be concluded from the studies described in this thesis. Seasonality in human fecundity was absent or negligible in countries such as the Netherlands with only moderate changes in the photoperiod, environmental temperature and seasonality in energy intake and energy expenditure. A seasonal pattern in poor reproductive outcome in such countries is unlikely to be the result of seasonality in fecundity, because of the negligible seasonal pattern in fecundity and the large proportion of conceptuses that are lost before a pregnancy is recognized. Human fecundity during IVF treatment is negatively associated with the woman's age. Moreover, it seems to be positively influenced by a prior gestation. If an IVF treatment has been performed without success, fecundity seems to be lower during the next treatment cycle in the case of idiopathic infertility in comparison with other indications for IVF treatment, and higher if prior embryo transfer has been performed. As the developed prognostic models which are described in this thesis did not always show good external validity, in general, they cannot yet be used for decision-making.

If further research into seasonality in fecundity is warranted, it should focus on populations who live under circumstances of extreme changes in the photoperiod or

temperature, or who show large seasonal variation in energy intake and energy expenditure. A theoretically-supported hypothesis on high-risk seasons is a prerequisite for studies on seasonality in poor reproductive outcome, because of seasonal variation in human fecundity. Consensus meetings concerning IVF treatment and prospective data collection on a national or even on an international basis are recommended. This will facilitate further research into, for instance, prognostic models for the cumulative probability of achieving a live birth after successive IVF treatments and models for predicting poor reproductive outcome. Moreover, several current IVF treatment modalities and future modalities should be evaluated and should focus not only on the effectiveness, but also on the effect on the development of subsequent offspring.

Samenvatting

Het vermogen van een vrouw om zwanger te worden, de fecunditeit, wordt uitgedrukt als de kans op zwangerschap per menstruatiecyclus bij vrouwen met zwangerschapswens. In dit proefschrift is onderzoek beschreven naar variatie in de fecunditeit onder natuurlijke omstandigheden en tijdens infertiliteitsbehandelingen. De kernvraagstellingen zijn: 1) Is er seizoenvariatie in de fecunditeit? en 2) Welke combinaties van factoren zijn voorspellend voor de resultaten van behandeling met in vitro fertilisatie (IVF)? Dergelijke kennis kan aanwijzingen geven voor preventieprogramma's ten aanzien van subfertiliteit en voor de toewijzing van infertiliteitsbehandelingen.

De hypothese dat er sprake is van seizoenvariatie in de fecunditeit bij de vrouw kwam voort uit het gegeven dat de meeste zoogdieren seizoengebonden voortplanting vertonen. Ook is seizoenvariatie waargenomen in diverse studies naar reproduktiefactoren bij de mens. Dit betreft bijvoorbeeld hormonen geproduceerd door de hypothalamus-hypofyse-ovarium as, ovulaties, spermakenmerken, zwangerschapskans, de wachttijd tot zwangerschap en de geboortefrequentie. Deze studies laten geen consistent patroon zien, zij het dat een zekere trend aanwezig lijkt te zijn naar een verhoogde fecunditeit van vrouwen gedurende het seizoen met lange dagen, terwijl de mannelijke vruchtbaarheid het beste lijkt te zijn gedurende het seizoen met korte dagen.

De eerste studies in dit proefschrift betreffen methodologische kwesties bij epidemiologische studies naar de voortplanting. Zo vormt de verscheidenheid in parameters een probleem bij het vergelijken van de resultaten van dergelijke studies. Drie veelgebruikte parameters zijn de prevalentie, de index en de ratio waargenomen versus verwacht. Aangetoond is dat de validiteit en de precisie van deze drie parameters hetzelfde zijn, maar dat het gebruikersgemak verschilt. In dit proefschrift is voor de prevalentie gekozen, omdat de statistische mogelijkheden van deze parameter de beste zijn (Hoofdstuk 2.1.1). Vervolgens is een toepassing van regressie-analyse gepresenteerd waarbij seizoenpatronen getoetst worden via een cosinusfunctie. Deze methode biedt de mogelijkheid voor verstorende factoren te corrigeren (Hoofdstuk 2.1.2). Tenslotte is via simulaties nagegaan of seizoenvariatie in zwangerschapsplanning een relatie tussen een seizoengebonden expositiefactor en ongunstige zwangerschapsuitkomsten kan verstoren. Het blijkt dat onder realistische omstandigheden een dergelijke invloed te verwaarlozen is (Hoofdstuk 2.1.3).

Seizoenvariatie in ovulaties is bestudeerd met behulp van gegevens van 407 vrouwen die in 1991-1992 voor het eerst naar de infertiliteitskliniek van het Academisch Ziekenhuis Nijmegen kwamen. Alleen vrouwen met menstruatiecycli korter dan 6 weken werden in dit onderzoek betrokken. Tijdens één cyclus werd nagegaan of er een ovulatie optrad door om de dag of dagelijks de follikelontwikkeling via transvaginale echoscopie te volgen en het progesteronniveau halverwege de luteale fase te bepalen. Er bleek geen seizoenvariatie in ovulaties te bestaan. Mogelijk is de variatie in daglengte, omgevingstemperatuur en energie-inname en -verbruik gedurende het jaar in Nederland te klein om tegenwoordig tot waarneembare seizoenvariatie in ovulaties te leiden, aangezien in twee andere studies bij populaties die leefden onder meer extreme omstandigheden wel seizoenvariatie in ovulaties geconstateerd was (Hoofdstuk 2.2).

Enige seizoenvariatie is gevonden in de resultaten van de eerste IVF-cycli van 1126 paren die in de jaren 1987-1993 behandeld zijn in het Academisch Ziekenhuis Nijmegen. Het beste resultaat trad op als de behandeling startte in de periode november-februari (Hoofdstuk 2.3). Het seizoen bezat echter geen voorspellende waarde toen alleen de gegevens van de eerste IVF-cycli in de jaren 1991-1994 (N = 757) gebruikt werden voor het ontwikkelen van prognostische modellen voor doorgaande zwangerschap (Hoofdstuk 3.2). Dit lijkt te wijzen op seizoengebonden veranderingen in behandelingsprotocollen in de beginjaren van IVF-behandeling in dit ziekenhuis.

Om seizoenvariatie in de wachttijd tot zwangerschap te bestuderen, is informatie van drie Deense studies gebruikt waarin zwangere vrouwen (N = 3.657) en werkneemsters in de textielindustrie en apothekersassistenten die zwanger waren geweest (respectievelijk N = 1.053 en N = 734) ondervraagd zijn over de wachttijd tot zwangerschap. Bij concepties in februari-april werd een hoge kans op een lange wachttijd waargenomen en bij concepties in augustus-oktober een lage kans. Dit patroon kan echter verstoord zijn geweest door seizoenvariatie in zwangerschapsplanning (Hoofdstuk 2.4.1). Daarom is een vervolgstudie uitgevoerd waarbij gebruik gemaakt is van een bevolkingsregister met informatie over 18.970 Franscanadese vrouwen die voor de eerste keer trouwden in de 17e of 18e eeuw. Als benadering voor de wachttijd tot zwangerschap is het interval tussen het huwelijk en de eerste geboorte minus 38 weken gebruikt. Er was sprake van een klein bimodaal seizoenpatroon met de meest fecunde perioden in december-januari en juni-juli. In deze studie is tevens geïllustreerd dat seizoenvariatie in de wachttijd tot zwangerschap sterk vertekend kan zijn als het seizoen gedefinieerd is aan de hand van de maand van conceptie terwijl er seizoenvariatie in zwangerschapsplanning is (Hoofdstuk 2.4.2).

Tenslotte is via literatuuronderzoek gezocht naar seizoenvariatie in het optreden van

Down syndroom. Ongeveer 70% van alle gevallen van Down syndroom is veroorzaakt door een maternale nondisjunctie gedurende de eerste meiotische deling. Deze deling vindt plaats gedurende de laatste dagen voor ovulatie. Seizoeninvloed op de fecunditeit zou dit stadium van de eicelrijping kunnen beïnvloeden door interactie met de hormonen die hierbij betrokken zijn. In een kritisch overzicht van twintig studies die gepubliceerd waren in de periode 1966-1995 werd geen consistent seizoengebonden patroon gevonden in de prevalentie van Down syndroom bij geboorte. Het is niet uitgesloten dat er wel een seizoenpatroon in Down syndroom op het moment van de bevruchting bestaat, maar verdwijnt door selectieve voortijdige vruchtdood. De beschikbare informatie was echter niet toereikend om dit te bestuderen (Hoofdstuk 2.5).

IVF-behandeling is bedoeld om de fecunditeit te verhogen. De in de literatuur gepresenteerde cumulatieve zwangerschapskans na IVF geeft echter een te optimistisch beeld van de bereikte fecundabiliteit na opeenvolgende IVF-cycli. Men neemt namelijk aan dat vrouwen die na een IVF-behandeling gestopt zijn zonder dat zij zwanger zijn geworden, bij een volgende IVF-cyclus dezelfde kans op zwangerschap zouden hebben als vrouwen die doorgaan met IVF. Omdat een slecht resultaat juist een veel voorkomende reden is om te stoppen met IVF-behandeling is er sprake van overschatting van de zwangerschapskans. Een realistischere benadering is dan ook aanbevolen waarin rekening gehouden wordt met de reden om met de behandeling te stoppen. Dit is geïllustreerd aan de hand van gegevens van 872 paren die voor de eerste keer in 1988-1992 in het Academisch Ziekenhuis Nijmegen behandeld zijn met IVF. De geschatte kans op doorgaande zwangerschap na vijf gestarte IVF-cycli was in deze groep 30% gegeven de meest behoudende veronderstellingen, 56% onder de optimistische assumptie die in de literatuur vaak gebruikt wordt en 34% als er rekening gehouden werd met de reden om te stoppen (Hoofdstuk 3.1).

Via prognostische modellen is nagegaan welke combinaties van patiëntkenmerken op een lage of hoge fecundabiliteit tijdens IVF-behandeling wijzen. Dergelijke modellen zouden gebruikt kunnen worden als hulpmiddel voor de gynaecoloog om patiënten te informeren over hun kans op succes met IVF-behandeling. Prognostische modellen voor de kans op doorgaande zwangerschap bij IVF-behandeling zijn ontwikkeld met behulp van gegevens van het Academisch Ziekenhuis Nijmegen uit 1991-1994 (N = 757). De modellen werden toegepast op data van het Catharina Ziekenhuis Eindhoven (N = 432). Bij aanvang van de IVF-behandeling bleken alleen de leeftijd van de vrouw en een voorgaande zwangerschap een voorspellende waarde te hebben voor doorgaande zwangerschap bij de eerste IVF-behandeling. Prognostische factoren voor de kans op

doorgaande zwangerschap bij de tweede IVF-behandeling waren: de leeftijd van de vrouw, onverklaarde infertiliteit en embryoterugplaatsing tijdens de eerste IVF-cyclus. Deze twee modellen voor de situatie bij respectievelijk het begin en het einde van de eerste IVF-cyclus vertoonden geen voorspellend vermogen in de populatie van Eindhoven. Een ander model voorspelde op het moment van de embryoterugplaatsing wel tamelijk goed de kans op doorgaande zwangerschap. Ofschoon het laatste model niet van klinische betekenis is, geeft het aan dat een mogelijke verklaring van het slechte voorspellend vermogen van de andere twee modellen zou kunnen liggen in verschillen in ovulatie-stimulatie protocollen en in de besluitvorming rond het stoppen van een behandeling om ovarium-hyperstimulatie syndroom te vermijden (Hoofdstuk 3.2). In een vervolgstudie (Hoofdstuk 3.3) is gevonden dat de twee modellen ook slecht voorspelden bij patiënten van het Academisch Ziekenhuis Nijmegen in 1995-1996 (N = 208) en van het Diaconessenhuis Voorburg (N = 1.424). Zelfs na kalibratie van de modellen door middel van een "krimpfactor" om een te sterke weerspiegeling aan de onderliggende gegevens te corrigeren, bleef de voorspellende waarde van de modellen teleurstellend. De resultaten van het testen van de modellen in de Nijmeegse populatie waren echter onnauwkeurig vanwege te weinig patiënten met een lage voorspelde kans. Als vrouwen in de Eindhovense patiëntenpopulatie bij wie de eerste IVF-cyclus afgebroken was vanwege de aanwezigheid van teveel grote follikels werden uitgesloten, dan bleek het mogelijk om de groep patiënten te identificeren die slechts ten hoogste 10% kans had op een doorgaande zwangerschap in de tweede IVF-cyclus. Het gaat hier om een derde van de patiëntenpopulatie. Dit is een mogelijke uitzondering op de algemene conclusie dat de prognostische modellen wegens gebrek aan externe validiteit nog niet in de klinische praktijk toegepast kunnen worden. Dat geldt niet alleen voor de in dit onderzoek ontwikkelde modellen, maar ook voor prognostische modellen die in de literatuur zijn gepresenteerd terwijl zij onvoldoende gevalideerd zijn. De kans op foute voorspellingen is doorgaans namelijk te groot.

Dat de fecunditeit van vrouwen afneemt bij het ouder worden is al lange tijd bekend. In een studie met gegevens van paren die een standaard IVF-behandeling ondergingen (N = 277) of een IVF-behandeling kregen met donoreicellen (N = 294) bij de Universiteit van Zuid-Californië in Los Angeles, Verenigde Staten, bleek een dergelijke afname ook waarneembaar te zijn bij IVF-behandeling met eigen eicellen, maar niet indien donoreicellen gebruikt werden. Veroudering van de baarmoeder lijkt dus niet van belang voor de afname van de fecunditeit. Een scherpe daling in de kans op doorgaande zwangerschap trad alleen op bij vrouwen die hun eigen eicellen gebruikten na de leeftijd van 40 jaar. Deze verminderde fecunditeit kon niet verklaard

worden door het kleinere aantal beschikbare eicellen en duidt op verminderde eicelkwaliteit bij deze oudere vrouwen (Hoofdstuk 3.4).

Uit de studies die in dit proefschrift beschreven zijn, kunnen de volgende conclusies getrokken worden. Seizoenvariatie in fecunditeit bij de mens lijkt niet of nauwelijks te bestaan in landen met een matige jaarlijkse variatie in daglengte, omgevingstemperatuur en energie-inname en -verbruik zoals Nederland. Dat een seizoenpatroon in ongunstige zwangerschapsuitkomsten in dergelijke landen veroorzaakt wordt door seizoenvariatie in de fecunditeit is onwaarschijnlijk, gezien de verwaarloosbare seizoenvariatie in fecunditeit en het verlies van een groot deel van de concepties voordat een zwangerschap geconstateerd wordt. De fecunditeit gedurende IVF-behandeling is negatief geassocieerd met de leeftijd van de vrouw. Het lijkt positief te worden beïnvloed door voorgaande zwangerschappen. Zodra een IVF-behandeling is uitgevoerd, lijkt de fecunditeit verlaagd bij onverklaarde infertiliteit als reden voor IVF-behandeling en verhoogd als tenminste één embryo is teruggeplaatst in de eerste cyclus. Aangezien de ontwikkelde prognostische modellen niet altijd extern valide waren, kunnen zij over het algemeen nog niet gebruikt worden bij besluitvorming.

Indien verder onderzoek naar seizoenvariatie in fecunditeit gewenst is, zou het gericht moeten zijn op populaties die leven onder extreme jaarlijkse variatie in daglengte, omgevingstemperatuur, energie-inname en -verbruik. Voor de studie naar seizoenvariatie in slechte zwangerschapsuitkomsten als gevolg van seizoenvariatie in de fecunditeit van de mens is een hypothese noodzakelijk waarin theoretisch goed onderbouwd is in welke seizoenen een hoog risico verwacht wordt. Consensusbijeenkomsten ten aanzien van IVF-behandeling en prospectieve dataverzameling op nationale of zelfs internationale basis worden aanbevolen. Dit kan verder onderzoek vereenvoudigen naar bijvoorbeeld prognostische modellen voor de cumulatieve kans op een levendgeboren kind na opeenvolgende IVF-behandelingen en naar modellen om de kans op ongunstige zwangerschapsuitkomsten te voorspellen. Het is noodzakelijk dat bestaande en nog te ontwikkelen varianten van IVF-behandelingen geëvalueerd worden op zowel hun effectiviteit als op de ontwikkeling van de nakomelingen, teneinde rationele toepassing mogelijk te maken.

Dankwoord

Promoveren lijkt een solistische aangelegenheid, maar dat is het allerminst. Vandaar dit woord van dank aan alle mensen die bij dit project betrokken zijn geweest, en in het bijzonder:

Prof. dr. ir. G.A. Zielhuis, beste Gerhard, heel veel dank voor je begeleiding. Mede doordat jij steeds met zo veel vaart je commentaar gaf, was het mogelijk om dit proefschrift in vier jaar af te krijgen. Ik ben blij dat je me de vrijheid gaf om buiten het keurslijf van deze AIO-functie te treden: eerst door de koerswijziging van seizoen naar IVF, later door mij de gelegenheid te geven in Amerika te werken.

Dr. C.J.C.M. Hamilton, beste Carl, mijn dank voor dat je me kennis hebt laten maken met het gynaecologische werkveld en voor je onderzoeksideeën. Mijn excuses dat je je soms opgejaagd voelde door mijn eeuwige "deadlines".

Prof. dr. A.L.M. Verbeek, beste André, met name dank voor je adviezen bij het prognostisch onderzoek en de discussies bij het refereren.

Prof. dr. F.J.M. Gabreëls, voor uw hulp bij de randvoorwaarden van dit project.

Dr. Piet Jongbloet, dit proefschrift is geënt op jouw idee. Ofschoon ik wat afgedwaald ben, hoop ik dat je het resultaat enigszins kan waarderen.

Drs. Huub Straatman, je statistische adviezen en programmeerwerk van enkele minder gangbare analyse-technieken waren me zeer welkom.

Collega's van de afdelingen Epidemiologie, Medische Statistiek, Informatiekunde en Technology Assessment, dank voor jullie gezelligheid en interesse. Astrid van Alst met name bedankt voor het verwerken van mijn vele e-mails om het contact tussen Amerika en Nederland te onderhouden. Erik Brummelkamp voor het tonen van allerlei trucjes om gegevensbestanden te koppelen. Hans Groenewoud voor hulp bij computerperikelen. Annelies Pellegrino voor de helpende hand bij statusonderzoek en gegevensinvoer. De stagiaires Francée Aarts, Miranda Reuvers en Eveline Geubbels voor hun inzet.

IVF-team Nijmegen, hartelijk dank dat ik de gegevens mocht gebruiken en voor de hulp om mij wat meer inzicht te geven in de praktijk van de IVF-behandeling. De andere IVF-centra: Dr. P.A. van Dop, Catharina Ziekenhuis Eindhoven en Dr. C.A.M. Jansen, Diaconessenhuis Voorburg, dank voor het beschikbaar stellen van de gegevens. Prof. dr. J.L.H. Evers, Academisch Ziekenhuis Maastricht, ook bedankt voor de gegevens, ofschoon ik ze niet kon benutten in mijn onderzoek. Prof. dr. E.R. te Velde,

Academisch Ziekenhuis Utrecht, dank voor uw advies bij het opstarten van het prognostisch onderzoek bij IVF.

Prof. R.J. Paulson, M.D., it was a pleasure to work at your IVF centre in Los Angeles, California, USA. Prof. dr. H.M. Vemer, Organon International B.V., Oss, and P. Weathersbee, Ph.D., Organon Inc., USA, thank you for introducing me to Prof. R.J. Paulson.

Prof. J. Olsen, Ph.D. M.D., J.L. Boldsen, Ph.D., and I. Schaumburg, Ph.D. M.D., thank you for giving me the possibility to work with the data of your studies and for your critical comments.

B. Desjardins, Programme de recherche en démographie historique of the University of Montreal, thank you for the data from the French-Canadian women in 17th and 18th century.

Familie en vrienden, dank voor jullie interesse in mijn bezigheden.

Roel, op de achtergrond was en ben jij steeds mijn klankbord. Mijn dankwoord aan jou op papier kan slechts een fletse afspiegeling zijn van de werkelijkheid.

Curriculum vitae

1965:	Geboren op 13 juni te Gouda.
1977 - 1983:	VWO, Sint Anthoniuscollege, Gouda.
1983 - 1989:	Voeding van de Mens, Landbouwuniversiteit Wageningen. Afstudeervakken: Voeding van de mens, epidemiologie, statistiek en voorlichtingskunde. Stage bij het International Agency for Research on Cancer, Lyon, Frankrijk.
1989 - 1992:	Onderzoeker, afdelingen Epidemiologie en Jeugdzorg, Gemeentelijke Gezondheidsdienst voor Rotterdam en omstreken.
1992 - 1996:	Assistent In Opleiding, vakgroep Medische Informatiekunde, Epidemiologie en Statistiek, Katholieke Universiteit Nijmegen. Onderzoek bij het centrum voor In Vitro Fertilisatie van de afdeling Gynaecologie en Obstetrie van de University of Southern California, Los Angeles, Verenigde Staten.
1996:	Docent Epidemiologie, vakgroep Medische Informatiekunde, Epidemiologie en Statistiek, Katholieke Universiteit Nijmegen.
1996-heden:	Onderzoeker, afdeling Obstetrie en Gynaecologie, Academisch Ziekenhuis Nijmegen.

Stellingen

behorende bij het proefschrift:

Human fecundity under natural conditions and during in vitro fertilization

I

Indien er geen sprake is van seizoenvariatie in de menselijke vruchtbaarheid, is hiervan geen seizoenvariatie in aangeboren afwijkingen te verwachten.

(dit proefschrift)

II

Als de externe validiteit van een prognostisch model niet aangetoond is, moet expliciet vermeld worden dat dit model niet voor praktische doeleinden bruikbaar is.

(dit proefschrift)

III

Een overkoepelende maat alleen is onvoldoende om de bruikbaarheid van een prognostisch model te beoordelen.

(dit proefschrift)

IV

Het prognostisch vermogen moet niet in sensitiviteit en specificiteit uitgedrukt worden, maar in de predictieve waarde van een positieve of negatieve test, aangezien het niet gaat om het herkennen van ziekte maar om het voorspellen wie ziek wordt.

(dit proefschrift)

V

De cumulatieve zwangerschapskans na opeenvolgende IVF-behandelingen neigt men te overschatten.

(dit proefschrift)

VI

Wie weet diep te zijn, streeft naar helderheid; wie voor de menigte diep wil schijnen, streeft naar duisterheid.

(F Nietzsche, De Vrolijke Wetenschap Amsterdam De Arbeiderspers, 1994)

VII

Reverence for what somebody said is a stultifying quality.

(W Somerset Maugham, Of Human Bondage London Penguin Books, 1992)

VIII

Wie gelooft, behoeft geen feiten.

IX

Zwangerschaps- en ouderschapsverlof werken bij tijdelijke arbeidscontracten niet emanciperend, maar discriminerend.

X

Seizoenen laten zien dat er meer is onder de zon.

Nijmegen, 6 februari 1997 A.M. Stolwijk