Can we distinguish between infertility and subfertility when predicting natural conception in couples with an unfulfilled child wish?

N. Van Geloven1,*, F. Van der Veen2, P.M.M. Bossuyt3, P.G. Hompes4, A.H. Zwinderman3, and B.W. Mol2

1Clinical Research Unit, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands 2Centre for Reproductive Medicine, Department of Obstetrics and Gynaecology, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands 3Department of Clinical Epidemiology, Biostatistics and Bio-informatics, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands 4Department of Reproductive Medicine, VU University Medical Centre, Amsterdam, The Netherlands

*Correspondence address. E-mail: n.vangeloven@amc.nl

Submitted on May 9, 2012; resubmitted on October 1, 2012; accepted on November 7, 2012

STUDY QUESTION: Can mixture survival models help distinguish infertility from subfertility in couples with an unexplained unfulfilled child wish?

SUMMARY ANSWER: Mixture models estimated that 47% of the couples were infertile; female age and previous pregnancy were significantly related to infertility, whereas duration of child wish was associated with a longer time to pregnancy for subfertile couples.

WHAT IS KNOWN ALREADY: To differentiate between couples who require assisted conception and couples who still have good chances of natural, i.e. unassisted, conception, several prediction models of natural conception have been developed. Prognostic factors in these models are usually assessed by Cox proportional hazard models that cannot differentiate between couples with an unfulfilled child wish who are subfertile, i.e. have reduced ability to conceive naturally, and couples who are really infertile, i.e. are completely unable to conceive naturally. We evaluated whether a mixture survival model can make such a distinction.

STUDY DESIGN, SIZE, DURATION: Consecutive couples presenting at the fertility clinics of 38 centres in the Netherlands between January 2002 and February 2004 joined a prospective cohort study. Of the 7860 couples in the cohort, 3917 couples met our inclusion criteria. The median follow-up was 219 days, with a maximum of 5 years.

PARTICIPANTS, SETTING, METHODS: Couples had to present with an unexplained cause of an unfulfilled child wish. A mixture model was used to estimate the proportion of couples who were infertile and the time to pregnancy for the subfertile couples.

MAIN RESULTS AND THE ROLE OF CHANCE: During the follow-up, 794 couples conceived naturally. The mixture model estimated that 47% [95% confidence interval (CI): 33–56%] of couples were infertile, despite the absence of objective factors indicating a cause for infertility. Of the evaluated prognostic factors, female age, duration of child wish, previous pregnancy, semen quality, BMI and cycle length, female age [odds ratio (OR): 1.11, 95% CI: 1.03–1.19] and previous pregnancy (0.22, 95% CI: 0.07–0.67) were significant predictors of infertility. Among subfertile couples, a longer duration of a child wish (FFR: 0.72, 95% CI: 0.61–0.85) was a significant prognostic factor for time to pregnancy. In the Cox models, all variables except BMI were significant predictors of time to pregnancy.

LIMITATIONS, REASONS FOR CAUTION: The mixture model had limited power due to a low number of couples at the end of the follow-up period. Mixture model analyses on external, long-term follow-up data are necessary to validate our results.

WIDER IMPLICATIONS OF THE FINDINGS: Mixture models could be a useful tool in selecting couples who require assisted reproductive technology because the effects of prognostic factors can be subdivided into effects on the fraction of infertile couples and effects on the time to pregnancy for subfertile couples, which is not possible in conventional models.
Introduction

Approximately 1 in every 10 couples with a child wish experiences involuntary non-conception (Hull et al., 1985; Gnoth et al., 2003; Wang et al., 2003). Given the risks of multiple pregnancy, costs and relatively unknown long-term side effects (Land and Evers, 2003; McLernon et al., 2010) of assisted reproductive technology (ART), gynaecologists have to be selective in offering ART to couples with an unfulfilled child wish and should advise expectant management when a couple has a reasonably good prognosis of a natural, i.e. unassisted, conception.

To make accurate predictions of a couple’s chance of conceiving naturally, several prediction models have been developed (Eimers et al., 1994; Collins et al., 1995; Snick et al., 1997; Hunault et al., 2004; Van der Steeg et al., 2007; Verhoeve et al., 2011). A recent review reported that seven of the nine identified studies on the prediction of natural pregnancy used Cox proportional hazard analysis to estimate the effect of patient characteristics on the probability of conception, while two used logistic regression (Leushuis et al., 2009).

In a logistic regression model, the binary outcome pregnancy has to be evaluated at a fixed time point, which requires all patients to have follow-up to at least that time point. Information from couples who were not followed up to the fixed time point cannot be used. Cox models have the advantage that information on couples with various lengths of follow-up can be combined. With these models one can estimate relative pregnancy rates conditional on couples’ characteristics such as female age, duration of unfulfilled child wish, previous pregnancy, body mass index, semen quality and others. A drawback of Cox models in predictions of natural conception is that they cannot differentiate between infertile couples, i.e. couples who are completely unable to conceive naturally, and subfertile couples, i.e. couples who are able to conceive naturally, but may take longer to do so. The ability to distinguish between these types of couples could lead to new management strategies. Couples who are expected to be infertile could be guided to undergo immediate ART, whereas subfertile couples with a good prognosis would benefit from initial expectant management.

Mixture survival models provide a way of modelling the time to an event of interest while accounting for a proportion of patients for whom this event will never take place. Such models have been frequently used in the analysis of cancer studies where the analyses have to account for cured patients who never experience the event of interest, relapse or disease-related death (Boag, 1949; Berkson and Gage, 1952; De Angelis et al., 1999). In that setting, mixture survival models are often referred to as cure models. In epidemiological applications, these types of models have been referred to as split-population models, referring to the split up of patients who are susceptible to the event of interest and patients who are not (Gray et al., 2010). In reproductive medicine, mixture models have been used in the past to study the success rates of ART (Guzick and Rock, 1981; Guzik et al., 1982, 1986; Peek et al., 1984; Guzik and Bross, 1992). The beta-geometric model introduced by Weinberg and Gladen, (1986) can also account for a subsample of infertile couples. This model was extended and has more recently been used to assess whether conception rates declining with age in the general population have to be primarily attributed to a larger proportion of couples absolutely unable to conceive in the older age groups or to a gradually decreasing monthly conception probability in most older couples (Wood et al., 1994; Dunson, 2000; McDonald and Rosina, 2001; Dunson et al., 2004).

The aim of this study was to assess whether mixture survival models can help to distinguish infertile couples from subfertile couples when predicting the probability of natural conception among a population of couples with an unexplained, unfulfilled child wish.

Materials and Methods

Patient selection

Our aim was to estimate the fraction of infertile couples and the effect of patient characteristics on this fraction and on the time to pregnancy in subfertile couples with a mixture survival model. In addition, we wanted to compare the results of such a mixture model with findings obtained from conventional Kaplan–Meier and Cox analyses. For these purposes, we used a data set collected in a cohort of couples presenting at the fertility clinics of 38 centres in the Netherlands between January 2002 and February 2004. Consecutive couples presenting at these clinics were invited to join a prospective cohort study. All consenting couples underwent a basic fertility workup according to the guidelines of the Dutch Society of Obstetrics and Gynaecology. The cohort study has been described in detail elsewhere (Van der Steeg et al., 2007, 2008).

Couples had to present with a regular ovulatory cycle, defined as a cycle length between 23 and 35 days with a cycle variation of ≤8 days. Ovulation was detected by a basal body temperature chart, mid-luteal serum progesterone or by ultrasonographic monitoring of the cycle. Women undergoing ovulation induction were excluded. Couples in whom semen analysis showed a severe impairment of semen quality requiring IVF–ICSI (defined as a post-wash total motile count <1×10⁶) were also excluded, as were couples with a history of reversal of sterilization, tubal surgery or IVF. Couples who had been referred by another gynaecologist (tertiary referral) and women who had been diagnosed with one or two-sided tubal pathology by hysterosalpingography, diagnostic laparoscopy or transvaginal hydrolaparoscopy were similarly excluded.

Follow-up

Women were followed from their first consultation at the fertility clinic until the occurrence of a natural ongoing pregnancy at 12 weeks of gestational age. Time to pregnancy was considered censored at the time when ART treatment (IUI or IVF) was started or at the last date of contact during follow-up, with a maximum of 5 years. We only included follow-up beyond a 1-year period of unfulfilled child wish. Couples for whom no follow-up information was collected after the first consultation were not included in the analyses.
Analysis
Because direct data on the proportion of infertile couples can never be obtained, it is necessary to use statistical models to study infertility fractions. We first estimated the fraction of infertile couples with a mixture model without including covariates. The model formulated the cumulative probability of pregnancy for the total patient group as \( (1 - \pi)^t F(t) \), where \( \pi \) is the fraction of infertile couples, \( (1 - \pi) \) is the fraction of subfertile couples and \( F(t) \) is the distribution function of the time to pregnancy for the subfertile couples. Given infinite follow-up time, all subfertile couples are expected to conceive and the cumulative pregnancy rate for the subfertile couples can never be obtained.

In a second step, we formed univariable models in which duration of child wish in years, female age in years (a spline approach with varying effect of age in years for different age categories did not significantly improve the fit of the mixture model), previous pregnancy in the current relationship, semen quality expressed as percentage progressive motility, body mass index (BMI) and cycle length in days were assumed to influence both the fraction of infertile couples (the logistic part of the model) and the time to pregnancy in subfertile couples, modelled through the exponential parameter of the model. Missing patient factors were imputed using a single imputation technique. Although in general a multiple imputation approach might better capture the uncertainty associated with unmeasured data values, we here chose to adhere to the approach used in previous analyses of the same data set (Van der Steeg, 2007).

In a third step, a multivariable mixture model was run, including all factors. The effects of the factors in the mixture models were expressed as odds ratios (OR), indicating the predictive effect on the fraction of infertile couples, and as fecundity rate ratios (FRRs) (hazard ratios) indicating the prognostic effect on fecundity in subfertile couples.

In a fourth step, we performed similar analyses using uni- and multivariable Cox proportional hazard models to compare the results of the mixture models with the findings obtained from conventional analyses of the same data. Here, the FRRs from the Cox models indicate the effect on fecundity for the total patient group, not just for the subfertile subgroup. The discriminative capacity of both the mixture and the Cox models was assessed with the c-index, representing the probability that a couple with a pregnancy at a certain time \( t \) has a higher calculated probability of pregnancy than a couple who has not yet achieved pregnancy at time \( t \). As the c-index is influenced by the number of censored observations and since in the mixture model the order of the calculated probabilities for two couples can reverse over time, we used a time-dependent variant of the c-index implemented in the pec package of the R statistical environment (Gerds et al., 2010; Gerds, 2012).

In a fifth step, we reassessed the contribution of the infertility component in the mixture model, now in the presence of covariates. The fit of the multivariable mixture model was compared with the fit of a multivariable exponential model containing the same patient factors but no infertility parameters, where the patient factors were all assumed to relate only to the fecundity in the total patient group. To this end, we used a likelihood ratio test with 7 degrees of freedom.

Results
Of the 7860 couples present in the cohort, 3917 couples met the criteria for inclusion in the analyses. The baseline characteristics of the

<table>
<thead>
<tr>
<th>Table I</th>
<th>Patient characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of couples</td>
<td>3917 (100%)</td>
</tr>
<tr>
<td>Female age (years)</td>
<td>32 ± 4</td>
</tr>
<tr>
<td>Duration of unfulfilled child wish (years)</td>
<td>1.5 (1–5.6)</td>
</tr>
<tr>
<td>Previous pregnancy in current relationship</td>
<td>1126 (29%)</td>
</tr>
<tr>
<td>Semen count (TMC × 10^6)</td>
<td>50 (2–379)</td>
</tr>
<tr>
<td>Semen count (% progressive motility)</td>
<td>41 (5–80)</td>
</tr>
<tr>
<td>Body mass index</td>
<td>23 (18–37)</td>
</tr>
<tr>
<td>Cycle length (days)</td>
<td>28 (24–33)</td>
</tr>
<tr>
<td>Follow-up duration (days)</td>
<td>219 (14–781)</td>
</tr>
<tr>
<td>Natural pregnancy during follow-up</td>
<td>794 (20%)</td>
</tr>
<tr>
<td>Time to pregnancy among conceiving patients (days)</td>
<td>138 (5–596)</td>
</tr>
</tbody>
</table>

Values are the mean ± standard deviation, median with 2.5th and 97.5th percentile or frequency with percentage.

Figure 1 Comparison of the cumulative pregnancy rate estimated with the Kaplan–Meier method and with the mixture model. The dashed red horizontal reference line represents the fraction of infertile couples as estimated by the mixture model. The solid red curve representing the estimates from the mixture model will never cross this (asymptotic) plateau.
The results of the univariable and multivariable mixture models are shown in Table II. Higher female age (OR: 1.11, 95% CI: 1.03–1.19, \( P = 0.005 \)) and a previous pregnancy in the current relationship (OR: 1.58, 95% CI: 1.37–1.83, \( P = 0.001 \)) were significant predictors of infertility. These ORs correspond to a marginal predicted infertility fraction of 27% at the age of 25, 51% at age 35 and 74% at age 45. For couples with a previous pregnancy in the current relationship, the marginal predicted infertility fraction was 21 versus 55% in couples without a previous pregnancy. For the prognosis of time to pregnancy in subfertile couples, only the duration of the unfulfilled child wish came out as a significant prognostic factor (FRR: 0.72 per year, 95% CI: 0.61–0.85, \( P < 0.001 \)). This effect may reflect a selection of patients with low (but not zero) monthly conception probability presenting with a longer duration of child wish.

To visualize the distinction between these two types of effects, we plotted the (multivariable) mixture model prediction curves for previous pregnancy in Fig. 2A and for short or long duration of child wish in Fig. 2B (split at the mean duration of 1.9 years). The corresponding Kaplan–Meier curves are added to these plots. In Fig. 2A, the asymptotes of the two mixture model prediction curves are far apart, since a previous pregnancy lowers the likelihood of infertility. In Fig. 2B, the two predicted lines flatten off to more or less the same level, with patients with a shorter duration of child wish reaching this plateau sooner than patients with a longer duration of child wish, indicating that the duration of child wish is the prognostic factor for the time to pregnancy in subfertile couples. Figure 2B also illustrates that the infertile fraction cannot be estimated from the final level of the Kaplan–Meier curve. In the short duration group, a single event at 1583 days causes an increase in the curve of over 20%.

In Table II (Cox model), the prognostic factors for time to pregnancy in all couples as estimated by the conventional Cox models...
Discussion

We used a mixture model to distinguish infertile couples from subfertile couples when calculating the probability of natural conception in couples with an unfulfilled child wish. With the mixture model, we estimated that 47% of the population was infertile and accounting for this fraction significantly improved model fit. These infertile couples are expected to have a prognostic profile similar to that of women with blocked tubes or women with anovulation, but they can at present not be identified clinically as such since they show no abnormal findings in the fertility work-up.

The mixture model that we propose disentangled the effect of patient factors on infertility from the effect on the time to pregnancy in subfertile couples. Whereas the Cox model could only indicate an effect of prognostic factors on the net fecundity in the total patient population, the mixture model made the distinction that female age and a previous pregnancy were related to infertility while a longer duration of child wish was associated with longer time to pregnancy for subfertile couples.

Two factors (semen quality and cycle length) that were identified as significant prognostic factors when using the Cox model had no significant effect on either component of the mixture model: not on the infertility fraction nor on the fecundity rate in subfertile couples. We believe that this can be attributed to a lack of power in the mixture model. Our mixture model estimates two effects (parameters) for each factor, whereas the Cox model only estimates one, leading to lower power and a possible split of the effect of a patient factor into two smaller effects. Additional factors explaining the lack of power are the low number of couples at the end of follow-up in our data and the fact that the plateau phase of the mixture model was not yet fully reached. It is known that estimates of mixture models can be unstable when the plateau phase has not yet been reached and when there are a small number of patients in the study toward the end of follow-up (Peek et al., 1984; De Angelis et al., 1999; Sposto, 2002; Lambert et al., 2007; Andersson et al., 2011). Therefore, our study should be viewed as a proof of principle study.
and a mixture analysis on external, long-term follow-up data is necessary to validate our results.

The mixture model that we propose here makes the distinction between infertile and subfertile couples at the start of follow-up. The model cannot distinguish infertility present already at the start of a woman’s reproductive life and infertility acquired before the start of follow-up, e.g. due to ageing or trauma. Also, the model does not incorporate the possibility of couples becoming infertile during the follow-up period (Sposto, 2002). Long-term follow-up is needed to reliably estimate the infertile fraction, but longer follow-up will make the problem of couples becoming infertile during follow-up more serious. This downside of expectant management will occur in particular for women with higher age, which may be the group that would need prediction of infertility the most. Wood et al. (1994) and Yashin et al. (1998) described models that include transitions from fecund to infertile status in the general population, but analytical issues caused by the large number of parameters that need to be estimated were reported.

The percentage of couples out of the general population that are absolutely unable to conceive is usually estimated to be low

---

**Table III** C-indices for the models.

<table>
<thead>
<tr>
<th>Time since first consultation</th>
<th>1 year</th>
<th>2 years</th>
<th>3 years</th>
<th>4 years</th>
<th>5 years</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Univariable mixture models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female age (years)</td>
<td>0.56</td>
<td>0.55</td>
<td>0.55</td>
<td>0.53</td>
<td>0.59</td>
</tr>
<tr>
<td>Duration of unfulfilled child wish (years)</td>
<td>0.57</td>
<td>0.56</td>
<td>0.57</td>
<td>0.58</td>
<td>0.60</td>
</tr>
<tr>
<td>Previous pregnancy in relationship</td>
<td>0.55</td>
<td>0.56</td>
<td>0.55</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>Semen count (% progressive motility)</td>
<td>0.56</td>
<td>0.55</td>
<td>0.56</td>
<td>0.54</td>
<td>0.54</td>
</tr>
<tr>
<td>Body mass index</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.52</td>
</tr>
<tr>
<td>Cycle length (days)</td>
<td>0.54</td>
<td>0.51</td>
<td>0.50</td>
<td>0.52</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>Multivariable mixture model</strong></td>
<td>0.63</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Univariable Cox models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female age (years)</td>
<td>0.56</td>
<td>0.55</td>
<td>0.55</td>
<td>0.53</td>
<td>0.59</td>
</tr>
<tr>
<td>Duration of unfulfilled child wish (years)</td>
<td>0.57</td>
<td>0.56</td>
<td>0.57</td>
<td>0.58</td>
<td>0.60</td>
</tr>
<tr>
<td>Previous pregnancy in relationship</td>
<td>0.55</td>
<td>0.56</td>
<td>0.55</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>Semen count (% progressive motility)</td>
<td>0.56</td>
<td>0.55</td>
<td>0.56</td>
<td>0.54</td>
<td>0.54</td>
</tr>
<tr>
<td>Body mass index</td>
<td>0.50</td>
<td>0.50</td>
<td>0.51</td>
<td>0.51</td>
<td>0.48</td>
</tr>
<tr>
<td>Cycle length (days)</td>
<td>0.54</td>
<td>0.51</td>
<td>0.50</td>
<td>0.51</td>
<td>0.57</td>
</tr>
<tr>
<td><strong>Multivariable Cox model</strong></td>
<td>0.63</td>
<td>0.62</td>
<td>0.62</td>
<td>0.63</td>
<td>0.67</td>
</tr>
</tbody>
</table>
at $\sim 2$–5% (Larsen, 2000; Gnoth et al., 2005). With roughly 90% of couples achieving a natural pregnancy within the first year of trying (Gnoth et al., 2005), our estimate of 42% of couples being infertile in our cohort of patients with reduced fertility (median duration of unfulfilled child wish of 1.5 years) and exclusion of couples with a clear cause of failing conception is plausible. Dunson et al. (2004) estimated that in the general population $\sim 1\%$ of couples are unable to conceive with no evidence of an increase of this percentage by age. Also Wood et al. (1994) did not find such an age effect. Those studies however only included women under the age of 40, which has likely obscured the age effects. The rather strong relation we found between age and infertility in our cohort with 129 women over the age of 40 is in line with McDonald and Rosina (2001), who found increased infertility for women aged 39–45, and with studies assessing infertility based on surveys in African countries (Larsen and Menken, 1991; Larsen, 2000) and with studies using historical (16–18th century) birth registries (Yan and Larsen, 2001; Leridon, 2004, 2008).

From the mixture model, one can retrieve prediction curves showing the probability of a couple being infertile based on the patient factors and the duration of follow-up. These curves could give valuable input in the decision to start ART therapies. For a couple with sufficient potential to ever conceive naturally the period of expectant management could be prolonged, whereas for a couple with a high probability of being infertile, immediate ART could be advised. In this way treatment resources can be allocated primarily to couples who are expected to never conceive naturally.

In our current analyses, couples who fell out of follow-up due to the start of ART were assumed to have a similar pregnancy prognosis to those who remained in follow-up. This so called non-informative censoring assumption is known to strongly influence predictions of natural pregnancy (Van Geloven, 2012). If a more conservative assumption was made for patients censored because of start of treatment, using competing risk models for example, the Cox models and the mixture models would yield different results. A mixture model that incorporates the competing risk of treatment might result in a more stable estimation of the fraction of infertile couples. This should be explored in future research.

We conclude that the mixture model could be a valuable tool in distinguishing infertile from subfertile couples and would thus help with the selection of couples who should be advised to start ART. However, data with sufficient long-term follow-up are needed to have sufficient power in the mixture approach.

### Authors’ roles

N.G., B.M., K.Z.: conceptualisation of manuscript. N.G.: analysis. F.V., P.M., P.H.: drafting the manuscript. All authors: critical revision of manuscript and approval of final manuscript.

### Funding

This study was supported by grant 945/12/002 from ZonMw, the Netherlands Organization for Health Research and Development, The Hague, the Netherlands.

### Conflict of interest

None declared.

### References


Distinguishing between infertility and subfertility


