Treatment should be considered a competing risk when predicting natural conception in subfertile women

N. Van Geloven¹*, K.A. Broeze², P.M.M. Bossuyt³, 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

*Correspondence address. E-mail: nanvangeloven@gmail.com

Submitted on June 23, 2011; resubmitted on November 26, 2011; accepted on December 1, 2011

BACKGROUND: Prediction of natural conception in subfertile couples can help to differentiate between couples who should have immediate treatment and couples who can aim for natural conception for some time. Natural conception rates are often estimated using standard techniques such as Kaplan–Meier or Cox proportional hazard models. These estimates can be biased by incorrect handling of data from women who start assisted reproductive technology therapy before the end of the follow-up period. This paper discusses the validity and the impact of the assumption of non-informative censoring as used in the Kaplan–Meier and Cox models.

METHODS: In a cohort of 5360 subfertile couples with suspected tubal pathology, the probability of natural conception and the prognostic value of additional tests for tubal pathology were estimated using traditional methods and with a competing risks analysis.

RESULTS: The estimated probability of natural conception within 3 years was almost 2-fold higher when assuming non-informative censoring compared with the competing risks model, 41 versus 22%. The prognostic value of tests was more conservative using the competing risks model than with the traditional methods, the fecundity rate ratio for Chlamydia antibody testing was 0.72 versus 0.67, for hysterosalpingography, 0.83 versus 0.71 and for diagnostic laparoscopy, 0.89 versus 0.74.

CONCLUSIONS: Given the improbable validity of the non-informative censoring assumption, the predictions of natural conception and of the prognostic value of tests are likely to be overestimated by the traditional analytic methods. We suggest the use of competing risks models as an alternative, more conservative, form of analysis when predicting natural conception and evaluating prognostic fertility tests.

Key words: natural conception / censoring / competing risks

Introduction

When subfertile couples present at a fertility clinic, basically two treatment strategies are at hand: expectant management or assisted reproductive technology (ART) therapy, such as IUI, IVF or ICSI. The choice between the two strategies is largely driven by the doctors’ expectation of the probability of natural conception.

Studies investigating the chances of a natural pregnancy try to quantify the probability that subfertile women will conceive given that no assisted treatment is started. These predictions are usually based on male and female patient characteristics collected during the fertility workup (Eimers et al., 1994; Collins et al., 1995; Snick et al., 1997; Hunault et al., 2004). A recent review by Leushuis et al. (2009) reported that seven out of the nine identified models on natural pregnancy used Cox proportional hazard analysis to estimate the effect of patient characteristics on the conception probability, while two studies used logistic regression (Bahamondes et al., 1994; Jedrzejczak et al., 2008).

With an estimated prevalence between 11 and 30% in subfertile populations, tubal pathology is an important cause of an unfulfilled wish for a child (Hull et al., 1985; Collins et al., 1995; Snick et al., 1997). As additional tests for tubal pathology can further improve pregnancy predictions, several studies have evaluated the three main procedures for tubal testing: Chlamydia antibody titre (CAT) tests,
underwent a basic fertility workup according to the guidelines of the Dutch centre gave Institutional Review Board approval for this study. All couples enrolled in a prospective cohort study. The local ethics committee of each participating centre approved the study. Between January 2002 and February 2004, consecutive couples presenting to the fertility clinic were included. The patient selection process is detailed elsewhere (Van der Steeg et al., 2007; Verhoeve et al., 2011). The patient selection process is summarized in Figure 1. The present analysis is limited to couples with a regular ovulatory cycle, defined as a cycle length between 23 and 35 days with a within-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. When ovulation induction therapy was used, despite an initial regular cycle, patients were also excluded from our analyses. Couples in whom semen analysis showed a severe impairment of semen quality requiring IVF–ICSI (defined as a post-wash total motile count <1×10^6) were excluded, as were couples with a history of reversal of sterilization, tubal surgery or IVF. After the basic fertility work-up, a prognosis for natural pregnancy was calculated, and was used to counsel couples for either immediate treatment or expectant management.

Follow up
Patients were followed from their first consultation at the fertility clinic. The follow-up period could be ended by one of the following events: (i) the occurrence of a natural ongoing pregnancy at 12 weeks of gestational age; (ii) the start of ART therapy (IUI or IVF); (iii) cancellation of study participation on the patient’s request while on expectant management; (iv) the end of the planned study period while on expectant management. Only in the first case, an exact observation of the time to natural pregnancy could be obtained; the other three situations led to censored observations of time to pregnancy. Couples for whom no follow-up information was collected after the first consultation were not included in the analyses.

Estimating the probability of natural pregnancy
We first estimated the probability of natural pregnancy by calculating Kaplan–Meier estimates (Kaplan and Meier, 1958). The follow-up period was restricted to a maximum of 3 years. For computational details on the Kaplan–Meier method, we refer to the Supplementary data. The Kaplan–Meier approach assumed that all censoring mechanisms were non-informative, or, stated differently, that patients who were still in follow up at a certain moment in time were representative of the total Society of Obstetrics and Gynaecology. This workup has been described in detail elsewhere (Van der Steeg et al., 2007; Verhoeve et al., 2011).

Materials and Methods

Patient selection
Between January 2002 and February 2004, consecutive couples presenting at the fertility clinic of 38 centres in The Netherlands were invited to join a prospective cohort study. The local ethics committee of each participating centre gave Institutional Review Board approval for this study. All couples underwent a basic fertility workup according to the guidelines of the Dutch
patient group. This would imply that patients for whom follow up was ended because they started on ART therapy had a probability of natural pregnancy similar to that of patients who remained in follow up.

To study whether the non-informative censoring was a reasonable assumption for patients who started treatment, we compared three known risk factors for subfertility between patients who were not referred for treatment and patients who started treatment: female age, spontaneous pregnancy in the current relationship and duration of subfertility, measured as the time between their wish for a child and the first consultation. Since patients with a favourable profile may not have started treatment (because they became pregnant before considering treatment), we also compared the risk factors of treated patients with patients who did not get pregnant or start treatment. We used t-test and a χ² test statistics for these comparisons.

We then estimated the probability of natural pregnancy by an alternative competing risks analysis. The competing risks model assumed non-informative censoring only for patients whose follow-up period ended while still on expectant management. The start of ART therapy was considered a competing risk for natural pregnancy as a patient can only experience one of these events. This implied that after treatment initiation, chances of a natural pregnancy were assumed to be zero. The computational details of the competing risks model are given in the Supplementary data. For a more comprehensive explanation of the method, we refer to Putter et al. (2007).

**Figure 2** Kaplan–Meier curves for (A) natural pregnancy; (B) IUI treatment and (C) IVF treatment.
assumed to have a zero probability of natural pregnancy from that time on. For details on the method, we again refer to Putter et al. (2007) and to Fine and Gray (1999).

For simplicity, we only studied dichotomous test results. The following CAT tests were used: ELISA (cut-off 1:1), Immunofluorescence (cut-off 1:32) and micro-immunofluorescence (cut-off 1:32) (Van der Steeg et al., 2008). The HSG and DLS results were classified as positive when at least one of the fallopian tubes showed an occlusion. The time to pregnancy was calculated from the day the test of interest was performed. As the timing of the CAT test was not recorded for almost half of the cases, we used the date of first consultation as the starting point for the CAT analysis. We studied a maximum follow-up period of 3 years for the CAT test, 2 years for the HSG test and 1 year for the DLS test. This way, we were sure to have sufficient patient numbers for the analyses in all three models. The number of patients included in the analyses depended on the number of patients exposed to the test at hand (n = 4281 CAT, n = 2271 HSG and n = 1164 DLS, Fig. 1).

Results

Probability of natural pregnancy

The 5360 patients in our cohort had a median follow-up duration of 225 days. During this time 1005 patients conceived naturally, 3012 started treatment (2290 IUI and 722 IVF) and the remaining 1343 were still on expectant management at the end of their follow up. The Kaplan–Meier estimates for the natural pregnancy rate resulted in the cumulative pregnancy (one-minus-survival) curve (Fig. 2A) and the probability of a natural pregnancy within 1 year was estimated as 23%. The probability of conception within 2 years was 36%; for the 3-year follow-up period this probability was 41%. For illustration, we also performed Kaplan–Meier analyses in which treatment was the event of interest. The curves in Fig. 2B and C show the estimated probabilities of receiving IUI and IVF treatment, respectively.

When we compared risk factors between treated patients and patients not referred for treatment in our data set, we observed that untreated patients had a more favourable profile compared with patients who started treatment: these patients were significantly younger (mean 32.1 versus 32.8 years; P < 0.001), more often had a previous spontaneous pregnancy in their current relationship (35 versus 27%; P < 0.001) and had a shorter duration of subfertility (mean 21.0 versus 25.3 months; P < 0.001). Although differences were smaller, the untreated patients who did not conceive also had a more favourable profile compared with patients who started treatment: untreated patients were younger (mean 32.5 versus 32.8; P = 0.02), more often had a previous spontaneous pregnancy in their current relationship (33 versus 27%; P < 0.001) and had a shorter duration of subfertility (mean 23.7 versus 25.3 months; P < 0.001).

The cumulative incidence curve for natural pregnancy resulting from a competing risks analysis was plotted in Fig. 3 (bottom line). The pregnancy curve was remarkably lower than the curve in Fig. 2A. One year after the first consultation, the cumulative pregnancy rate was estimated as 18%, after 2 years, as 22% and after 3 years of follow up, as 22%. Note that the estimated probability of natural pregnancy within 3 years in the competing risks setting was almost half that in the Kaplan–Meier analysis. We also plotted the estimated cumulative treatment incidences of IUI and IVF in the figure as surfaces. These estimates of the probability of treatment were also lower in the competing risk model compared with the Kaplan–Meier analyses. As a consequence of the assumptions made in a competing risks analysis, the summed cumulative incidences of natural pregnancy, IUI treatment and IVF treatment remained below 100% during the whole follow-up period. The estimate of natural pregnancy in the competing risks setting is an estimate for the probability of conceiving naturally, knowing that one may start treatment, after which the chances of a natural pregnancy were assumed to be zero. In the Kaplan–Meier setting the chance of a natural pregnancy after treatment initiation was similar to that of patients who did not start treatment, as did the Kaplan–Meier analysis. The competing risks models estimated pregnancy probabilities while assuming a zero probability of spontaneous pregnancy after treatment starts, explaining the lower level. Second, the curves for positive and negative test results in the right panel were closer to each other than in the left panel. Accordingly, the FFRs from the Fine and Gray models were closer to one than the FFRs from the Cox models: for the CAT test, the Fine and Gray model estimated an FFR of 0.72 [95% confidence interval (CI) 0.61–0.85], whereas the Cox model estimated an FFR of 0.67 (95% CI
0.57–0.79); for the HSG the Fine and Gray model estimated an FFR of 0.83 (95% CI 0.63–1.08), whereas the Cox model estimated an FFR of 0.71 (95% CI 0.53–0.93); for the DLS the Fine and Gray model estimated an FFR of 0.89 (95% CI 0.53–1.30), whereas the Cox model estimated an FFR of 0.74 (95% CI 0.47–1.18). This means that the Cox models estimated a stronger effect of test results on pregnancy prognosis than the competing risks models.

Discussion

Our analyses showed that the non-informative censoring assumption used in the Kaplan–Meier and Cox model has a strong effect on the estimates of natural pregnancy rates and on the prognostic ability of tests. Estimated pregnancy rates were up to 2-fold higher and a stronger FFR was estimated for all three evaluated tests compared with that found with the competing risks analyses that do not assume non-informative censoring for patients who started ART therapy. A comparison of known risk factors for fertility showed that untreated patients had a more favourable profile than patients who started treatment, which contradicts the non-informative censoring assumption. The difference between the estimated pregnancy rates in both methods was larger at later follow-up times, as the number of treated patients and the difference in assumed pregnancy rates for treated patients increased with time.

From the analysis of this cohort of subfertile patients, we wanted to learn the probability of a natural conception for patients in whom treatment is never started. Such a probability would give the best input for the treatment decision of new patients. The question is whether the traditional approaches give a valid estimate of this treatment independent pregnancy probability. As we only had data obtained from the cohort study, where patients did start treatment, the traditional estimators used the firm assumption that the prognosis of patients in whom treatment was started was, from that point on, similar to the prognosis of untreated patients still in follow up. The probability of natural pregnancy was very likely overestimated by the Kaplan–Meier and Cox models, since untreated patients who stayed in follow up for spontaneous pregnancy probably had a better prognosis than patients who started treatment. This was confirmed by a comparison of risk factors between treated and untreated patients. The overestimation is expected to be most profound at later time points, where the observed pregnancy rate of only a small number of patients at risk was extrapolated to a large number of patients.

Figure 4 Comparison of Cox models (left plots) and Fine and Gray models (right plots) for CAT − (n = 3005) versus CAT + (n = 1276), HSG − (n = 1868) versus HSG + (n = 403) and DLS − (n = 866) versus DLS + (n = 298).
who had already started treatment. The prognostic ability of diagnostic tests could have also been overestimated by the traditional models, since not only patients with a worse prognosis, but also patients with tubal tests indicative for tubal pathology, had a higher probability of starting ART therapy. Both mechanisms jointly caused a selection of test negative patients with relatively good prognosis in later follow up. In our data, this selection may have led to an overestimation of the prognostic ability of the diagnostic tests in the Cox model.

We presented an alternative analysis using competing risks techniques. This approach modelled the occurrence of pregnancy and treatment jointly, and used the assumption that one cannot conceive naturally after start of treatment. When estimates from these models are used for answering the important clinical question ‘what would the prognosis for a new patient be if treatment is never started’, we basically assume that patients for whom treatment was started in the past from that moment on had zero probability of a natural pregnancy (even in the hypothetical case where treatment would not have been started). However, the differences in known risk factors found between treated and untreated patients were, although statistically significant, modest. Therefore, the assumption of zero pregnancy probability after start of treatment could have been too conservative, possibly leading to an underestimation of the pregnancy rates by the competing risk models.

It is clear however, that the sole use of traditional Kaplan–Meier and Cox models is not satisfactory for modelling time-to-natural-pregnancy data. A competing risks approach might better suit reality. Simultaneous use of both models could provide an optimistic and a conservative estimate for the natural pregnancy probability. Although such a scenario approach might complicate treatment decisions, it does justice to the fact that we are trying to estimate treatment-independent pregnancies from patients who actually received treatment, and for whom we have no concluding evidence on natural pregnancy prognosis.

The problem of informative censoring in the prediction of natural pregnancy has been described before (Collins et al., 1995; Mol et al., 1997). Collins et al. (1995) compared the use of untreated couples only in the prediction of natural pregnancy to the use of both untreated and treated persons, terminating observation for the latter at the time of the first treatment. They advocated the second approach, arguing that using only untreated couples would overestimate conception prognosis. In our opinion, prognosis is still overestimated when analysing untreated and treated patients together in a model that assumes non-informative censoring, as untreated patients are overrepresented at later time points. Mol et al. (1997) performed a sensitivity analysis by excluding patients with a prognostic factor possibly involved in informative censoring, factors prognostic for both pregnancy and treatment initiation. An analysis of patient factors that are prognostic for both pregnancy and treatment initiation can be an auxiliary step in exploring whether data might be subject to informative censoring. However, excluding patients with such prognostic factors leads to loss of data and makes the results of the analysis less generalizable.

Hunault et al. (2004) stated that data from several decades ago are better suited to analyse natural pregnancy than current data, as treatment intensity was lower then. Indeed, the joint data set used in their influential paper showed a lower raw treatment rate (22% at 1 year follow up) than the more recent data set used for our analyses (42% at 1 year follow up). It would be interesting to see how the widely used predictions from Hunault’s model change when using the more conservative competing risks approach.

The conclusion drawn from this manuscript that conventional methods overestimate the natural pregnancy rates has consequences for previous reports on this subject (Mol et al., 1999; Van der Steeg et al., 2007; Verhoeve et al., 2011). Van der Steeg et al. (2007) evaluated the calibration of their model by comparing the predicted versus observed pregnancy rate, but as the observed pregnancy rate was again estimated assuming non-informative censoring, overestimation could not be noted. However, we want to stress that before the introduction of prognostic models in clinical practice, the utility of these models should have been established in impact studies (Leushuis et al., 2009). As impact studies have shown that the use of prognostic models reduces costs without compromising pregnancy rates, we feel that the current prognostic models are useful in clinical practice, despite the fact that during model building, pregnancy rates were overestimated (Steures et al., 2006).

Ideally, data on the natural fertility course would be collected in the absence of treatment. Because patients demand treatment for subfertility, such data will never become available for a long follow-up period. Consequently, the estimation of pregnancy probabilities and prognostic value of patient factors will always be influenced by therapeutic interventions (Mol et al., 1997). Therefore, we conclude with a strong recommendation for a careful evaluation of the assumptions made for patients who start treatment and we advocate the use of competing risks models when analysing natural pregnancy rates from cohort data whenever couples fall out of follow up because of the start of ART therapy. This way we can generate more realistic information to counsel patients, evaluate the prognostic capacity of diagnostic tests and guide the formulation of policies for subfertility management and resource allocation.

Supplementary data
Supplementary data are available at http://humrep.oxfordjournals.org/.

Authors’ roles
N.V.G., B.W.M., A.H.Z. and P.M.M.B. were involved in conceptualization of manuscript. N.V.G. performed the analyses. K.A.B. drafted the manuscript. All the authors critically revised the manuscript and approved the 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.


