Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples

Sir,

Models are nowadays being put forward as an essential tool for clinical decision-making and in counselling patients. They are introduced as a necessary tool to prevent unnecessary treatment. In fact that is what a clinician wants: a crystal ball to predict pregnancy in order to counsel patients in a reliable way. Previously, in a paper by Hunault et al. (2004), a synthesis model was introduced based upon the original data of three previously published studies to predict spontaneous pregnancy in a subfertile population (Eimers et al., 1994; Collins et al. 1995; Snick et al., 1997).

In the recent paper by Van der Steeg et al. (2007), the dream finally seems to come true, since a prospective validation study resulted in a paper bearing the title ‘pregnancy is predictable…’. A prospective validation study is described, in which a consecutive cohort of 3021 subfertile couples was studied. Our centre was among the 38 centres that included patients. An essential step in the validation of the model of Hunault et al. (2004) was performed: the authors validated the model by applying it to another population than the one in which it was developed. This is important because prediction models tend to be overoptimistic when applied in the population in which they are developed compared with when applied in other populations (Stolwijk et al., 1996), although the authors state differently in the manuscript.

From the calibration plot in the paper, where the cohort was split up into 10 groups based on the deciles of the calculated probabilities, it becomes clear that the majority of the patients has a predicted probability which is around the mean predicted cumulative ongoing pregnancy rate (32%). As this percentage will be the prior probability of a spontaneous pregnancy without any information about the patient, we feel that a prognostic model is only useful if it changes this prior probability in an accurate way.

Furthermore, it was stated that in 1104 couples (36%) the probability of a spontaneous pregnancy was over 40%, being the cut-off percentage for expectant management. In the calibration plot, however, only two deciles corresponding with 611 couples with a probability of over 40% were described. This leads to speculation about the remaining couples.

Lastly, by validating the model that was also used in the decision whether or not to treat patients, a bias towards better calibration was introduced.

In contrast with the authors, we conclude from the paper that for most patients pregnancy cannot be predicted in a reliable way. The discriminative capacity, expressed by the c-index (which is comparable to the area under the ROC-curve) was reported to be 0.59. (One has to remember that a perfect model has a c-index of 1.0 and that a c-index of 0.5 means that prediction is no better than the flip of a coin.)

Previously, another model was validated by our group and since a c-index of 0.63 was found, we concluded that for the majority of the women the model did not give any more certainty (Smeenk et al. 2000).

Van der Steeg et al. (2007) state in the discussion that it is more important to accurately estimate whether a couple has a high or low chance to conceive than to know exactly which couple will conceive and which will not. However, this is the problem in counselling, as the couple wants an individual approach. In fact, the couple does want to know whether they will become pregnant or not and a perfect model would indicate a pregnancy chance of 0 or 100%. The high percentage of couples being treated within 6 months after the workup, sometimes despite their high prognosis and expectant study design, underlines this issue.

In summary, models to predict pregnancy can well be used to identify different prognostic groups in a population and thereby serve as a tool in patient counselling and shared decision-making. The discriminative capacity however is very limited, meaning that models provide limited information about who will or will not be pregnant.

References


Reply: Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples

Sir,

We thank Drs Smeenk and colleagues for their interest in our work. The authors address several issues. First, they state that a prognostic model is only useful if it changes the prior probability for pregnancy in an accurate way, and suggest that the Hunault model does not change the prior probability of 32% that most couples would have for spontaneous conception within 12 months. We disagree with the interpretation of Smeenk et al. Figure 3 of our article shows clearly that the 10% of the couples with the best prognosis have a mean predicted probability of over 50%, and indeed 50% of these couples conceive. Even so, of the 20% of couples that have a predicted probability of over 40%, two deciles corresponding with 611 couples were based on the model without the PCT.

Secondly, Smeenk et al. remark that in 1104 couples (36%) the probability of a spontaneous pregnancy was over 40%, whereas only two deciles corresponding with 611 couples with a probability of over 40% were described in the calibration plot. The authors are correct to point this out, which provides us with the opportunity to clarify this apparent discrepancy. The number of 1104 couples was based on predictions with a model with the Post Coital Test (PCT), whereas the 611 couples were based on the model without the PCT.

Finally, the authors claim that our work was threatened by bias because the evaluated model was also used to decide which couple would be treated and which couple not. This criticism would be true if we had not considered the time that each couple was trying to conceive without medical assistance. However, as we used time to spontaneous conception as outcome measure, we took into account that some couples were treated soon, whereas others waited longer before treatment was started.

Smeenk et al. then assess the prediction model with the c-index and go on to state that pregnancy cannot be predicted in a reliable way with a poor discriminative capacity. They state that a model in order to be useful has a virtually perfect discriminative capacity (c-index or area under the curve, AUC of 1.0). Prognostic tests differ from diagnostic tests concerning their discriminative capacity. Diagnostic tests, for example a laparoscopy for the diagnosis of two-sided tubal pathology, often show a good discriminative capacity with a high AUC. Prognostic tests have an inherent level of difficulty that differs from problem to problem. In reproductive medicine, prognostic models show AUCs around 0.6, when validated in an external population. In reproductive medicine the consequence of a false-negative prediction, i.e. the prediction that pregnancy will occur whereas it turns out that the couples does not conceive, only implicates a delay of treatment of 6 or 12 months. Especially for this good prognosis group, treatment will offer no benefit over an expectant management and as such the model helps us to offer tailored management without claiming to become fortune-tellers. As a consequence, the discriminative capacity (c-index) is often misinterpreted in reproductive medicine, as has been addressed in the longer and the recent past (Diamond, 1992; Mol et al., 2005). More important is the aim to create a model with a good calibration, in which we have succeeded. We have shown that prediction models can be used to classify subfertile couples accordingly, and we have shown that such an approach can prevent overtreatment (Steures et al., 2006) (http://www.freya.nl/probability.php).

In summary, we are not dreaming of crystal balls to make predictions in individual couples. We do feel however that it is important to identify subfertile couples who have a good or a poor prognosis for spontaneous pregnancy. We also feel that we have shown in our research that this validated model, albeit not a perfect model, goes a long way in this direction.

References


Jan Willem van der Steeg1,2,4,8, Pieteren Steures1,2,4, Marinus J.C. Eijkemans1, J. Dik F. Habbema1, Peter G.A. Hompes4, Frank J. Broekmans5, Thierry H.J.H.M. van Dessel6, Patrick M.M. Bossuyt7, Fulco van der Veen2 and Ben W.J. Mol2,7

1Department of Public Health, Erasmus MC, University of Medical Center, Rotterdam, 2Academic Medical Center, Center for Reproductive Medicine, Room H4-213, Department of Obstetrics and Gynaecology, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands, 3Department of Obstetrics and Gynaecology, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands.

1Correspondence address. Tel.: +31-24-361-4748; Fax: +31-24-366-8597; E-mail: j.smeenk@obgyn.umcn.nl

doi:10.1093/humrep/dem160
Advance Access publication on June 21, 2007