Clinical prediction models to inform individualized decision making in subfertile couples: a stratified medicine approach

DJ McLernon[1], ER te Velde[2], EW Steyerberg[2], BWJ Mol[3], S Bhattacharya[1]

[1]Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK; [2]Department of Public Health, Erasmus MC - University Medical Centre Rotterdam, Rotterdam, The Netherlands; [3]Department of Obstetrics and Gynaecology, Academic Medical Centre, Amsterdam, The Netherlands.

Corresponding author:

David J. McLernon, Medical Statistics Team, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, UK, AB25 2ZD.

Tel: +441224437152, Fax: +441224437285

Email: d.mclernon@abdn.ac.uk

Running title: Individualized decision making in subfertility

Word Count: 4139

Abstract

Infertility is defined as failure to conceive after one year of unprotected intercourse. This dichotomisation into fertile versus infertile, based on lack of conception over 12 month period, is fundamentally flawed. Time to conception is strongly influenced by factors such as female age and whilst a minority of couples have absolute infertility (sterility), many are able to conceive without intervention but may take longer to do so, reflecting the degree of subfertility. This natural variability in time to conception means that subfertility reflects a prognosis rather than a diagnosis. Current clinical prediction models in fertility only provide individualised estimates of the probability of either treatment independent pregnancy or treatment dependent pregnancy, but do not take account of both. Together, prognostic factors which are able to predict natural pregnancy and predictive factors of response to treatment would be required to estimate the absolute increase in pregnancy chances with treatment. This stratified medicine approach would be appropriate for facilitating personalised decision-making concerning whether or not to treat subfertile patients. Published models are thus far of little value for decisions regarding when to initiate treatment in patients who undergo a period of, ultimately unsuccessful, expectant management. We submit that a dynamic prediction approach, which estimates the change in subfertility prognosis over the course of follow-up, would be ideally suited to inform when the commencement of treatment would be most beneficial in those undergoing expectant management. Further research needs to be undertaken to identify treatment predictive factors and to identify or create databases to allow these approaches to be explored. In the interim, the most feasible approach is to use a combination of previously published clinical prediction models.

Introduction

Infertility is defined as “a disease of the reproductive system defined by the failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse” according to the World Health Organisation (WHO) and International Committee for Monitoring Assisted Reproductive Technology (ICMART) (Zegers-Hochschild et al., 2009). Absence of pregnancy within this time-period is interpreted as evidence of sterility by many couples, who then request immediate treatment.

In fact, the probability of conceiving is highly variable (te Velde et al. 2000) and genuine unresolved infertility or sterility, occurs in a minority (3–5%) of all couples (Greenhall and Vessey 1990). As couples who are more “fertile” tend to conceive early, the length of time couples have been unsuccessful at conceiving reflects the degree of subfertility. The term “infertility” is often used interchangeably with “subfertility” ((Gnoth et al. 2005, Gurunath et al. 2011, Habbema et al. 2004)). However, in this article we define as subfertile those couples in whom routine investigations have not been able to identify any absolute barriers to conception such as blocked Fallopian tubes, anovulation or azoospermia. Many of these couples are advised to undergo a period of expectant management, meaning that they continue trying to conceive naturally for a specified period of time before being offered treatment.

Data from non-contracepting populations (Bongaarts 1975) show that increase in the duration of unsuccessful unprotected intercourse is associated with decreasing chances of pregnancy. However, the definition of infertility as a failure to conceive within a year represents an oversimplification, as many couples in this group will conceive beyond one year (Bongaarts 1975, Snick et al. 1997). The only certain way of ‘diagnosing’ absolute infertility in subfertile couples, i.e. establishing with certainty that a couple is sterile, is lack of conception in women at the end of reproductive life. By then of course it is too late to rectify the situation by medical means. Thus, in order to be a clinically useful entity, subfertility needs to reflect the prognosis of a couple in terms of their ability to

conceive unaided. Such an approach recognises the fact that apart from duration, a woman's ability to conceive also declines with her age and depends on many other factors that vary the chances of conception such as frequency of intercourse, semen quality and pelvic pathology (Evers 2002).

Having acknowledged that subfertility represents a prognosis rather than an absolute diagnosis, it is worth considering the best way of assessing the chances of pregnancy for the purposes of initiating investigations and treatment. One option, which allows consideration of time on a continuous scale (rather than dichotomously) and a couple's risk factors for conception, is to use appropriately developed and validated clinical prediction models. Many of these already exist in fertility and they either predict the chances of pregnancy following treatment or without treatment, but not both (Leushuis et al. 2009). However, a method of taking both groups into account to estimate the additional chances of pregnancy following treatment could allow clinicians to identify those who would benefit from it. For example, an absolute increase of 5% in the chance of pregnancy following in vitro fertilization (IVF) compared to no treatment, might be important to a woman aged 38 whose natural chances of pregnancy have declined with age, but not to a woman aged 28 whose natural chances are still relatively high.

In this paper we describe the limitations of current clinical prediction models for subfertility and subsequently aim to explore the advancement of such models to address two key questions in fertility care: firstly, how should clinicians discriminate between those who need active fertility treatment versus those who do not? Secondly, given that subfertility prognosis changes over time, when should those on expectant management be offered active treatment?

Existing clinical prediction models in subfertility

Critical for the management of a subfertile couple prior to initiation of treatment is knowledge of their subfertility prognosis i.e. chances of spontaneous conception. As mentioned earlier, a way of

estimating subfertility prognosis is through clinical prediction modelling. A time-to-event statistical model (such as the Cox proportional hazards model) is a good method of predicting the chances of a binary outcome, such as conception (versus no conception), over a period of time. Such models adjust for prognostic factors, which are clinical or biological characteristics (such as female age and duration of infertility) that are associated with a clinical outcome (such as spontaneous pregnancy) in an untreated patient (Italiano 2011). Prognostic factors for subfertility can be obtained from the medical literature, clinical opinion or further research. Table 1 contains a list of known prognostic factors of spontaneous pregnancy from published models (Leushuis et al. 2009). The recently published Prognosis Research Strategy (PROGRESS) articles specify a framework of four interlinked themes for prognostic research. They recommend that large, prospective, registered prognostic factor studies with appropriate sample size and statistical analyses are required in order to find new prognostic factors that can predict an outcome (Hemingway et al. 2013, Hingorani et al. 2013, Riley et al. 2013, Steyerberg et al. 2013).

A systematic review of clinical prediction models in reproductive medicine identified 29 prediction models that predicted spontaneous pregnancy (n=9) or successful intrauterine insemination (IUI, n=3) or IVF n=17) (Leushuis et al. 2009). Of these 29 models, only eight were externally validated, three of which showed adequate performance (Custers et al. 2007, Hunault et al. 2004, Smeenk et al. 2000, Steures et al. 2006, Steures et al. 2004, Templeton et al. 1996, van der Steeg et al. 2007). Assessment of the predictive ability and external validation of a prediction model is essential if it is to be used to facilitate clinical practice (Collins 2005, Coppus et al. 2009). Aspects to evaluate include discrimination (how good a model is distinguishing between patients who do and do not become pregnant) and calibration (agreement between the probability estimate from the prediction model and observed outcome frequencies) (Steyerberg 2009).

The Hunault model, synthesised from three previous models based on three prospective databases of subfertile women attending a Dutch University hospital, a Dutch general hospital and eleven

Canadian University Hospitals, was found to predict spontaneous pregnancy leading to live birth reasonably well (Hunault et al. 2004). It had poor discriminatory ability, which is generally the case with prediction modelling in subfertile couples who tend to be rather homogeneous in terms of clinical characteristics (Coppus et al. 2009), but calibrated well when applied to external cohorts (Hunault et al. 2005, van der Steeg et al. 2007).

Two other models, which showed acceptable performance in the Leushuis et al (2009) review, were the Steures et al (2006) model which predicts live birth following IUI, and the Templeton et al (1996) model which predicts live birth following IVF. Both of these models also had poor discriminatory ability (Coppus et al. 2009, Smeenk et al. 2007). However, the Templeton model performs reasonably well after adjusting for improved IVF success rates over time (te Velde et al. 2014). Since the Leushuis review, a model developed using the Human Fertilisation and Embryological Authority (HFEA) database of all IVF treatments in the UK has been published (Nelson and Lawlor 2011) but performed no better than the Templeton model (te Velde et al. 2014).

A prognostic model could be used to make risk-based decisions in clinical practice. This would involve calculating the absolute chance of spontaneous pregnancy occurring within a pre-specified time period, e.g. one year, for a given individual (see Figure 1, Model 1a). Decisions regarding whether or not to treat can then be made using some pre-specified clinically agreed chance cut-off. For example, the creators of the Hunault model considered couples with <20% chance of spontaneous pregnancy as a poor prognosis group who should undergo immediate treatment (Hunault et al. 2004). Those with >40% chance were labelled as having a high chance of spontaneous pregnancy and the article suggested that these couples should be encouraged to wait for another year. Those in the middle group of 20–40% chance should be advised in such a manner as to balance the probability of pregnancy against the risks from fertility treatment.

However, using probabilities from a model that predicts treatment independent pregnancy to make treatment decisions does not take into account the chance that treatment may not be effective in

particular women. For example, being led solely by the above model cut-offs, a woman with a 15% chance of pregnancy would undergo immediate treatment. However, depending on the woman's specific characteristics, her chance of pregnancy following treatment may be no greater, or, it may be substantially greater. Conversely, models that predict pregnancy following treatment do not tell us whether the woman's absolute chance of pregnancy would have been any lower without treatment, and indeed how much lower (Figure 1, Model 1b). The best option would be to use a combined dataset, ideally from randomized controlled trial (RCT) data, including these two groups of women in order to model the additional benefit of treatment over no treatment. This can be made possible using a stratified medicine approach.

Absolute versus relative risk

Before we consider stratified medicine it is important to define absolute and relative risk. Absolute risk refers to the chance that a patient will have some outcome of interest (for example, a treated patient has a 10% risk of mortality and a control patient has a 12.5% risk of mortality). The relative risk refers to the chance of the outcome for one group of patients compared with another (in the given example the relative risk of mortality decreases by 20% for the treatment group compared to the control group). The word 'risk' is used since the outcome is often unfavourable. However, since pregnancy is a favourable outcome the term 'risk' is generally replaced with 'chance'. If the relative effect of treatment is constant for all patients, then the absolute benefit of treatment only increases in relation to the baseline pregnancy chances. For example, if statins have a constant relative risk reduction for all, then the absolute benefit is highest for those at highest risk of cardiovascular disease (LaRosa *et al.*, 1999).

To treat or not to treat? - A stratified medicine approach

Stratified medicine has been defined as 'the targeting of treatments (including pharmaceutical and non-pharmaceutical interventions) according to the biological or risk characteristics shared by subgroups of patients' (Hingorani et al. 2013). A clinician will use such an approach where the relative effect of treatment is believed to be inconsistent across patients. This means one or more patient characteristics are associated with changes in the relative effect of treatment. Such characteristics are called predictive factors of treatment response (Hingorani et al. 2013). The stratified medicine approach allows targeting of therapy based on the combination of subfertility prognostic factors and such treatment predictive factors, which increase the response to treatment in relation to no treatment. This enables decisions to be made regarding who should receive such treatment. For example, in non-small cell lung cancer, the response of the disease to chemotherapy is quite poor but there are therapy agents, gefitinib and erlotinib, which optimise therapy by being effective only in patients whose tumours harbour specific epidermal growth factor receptor profiles (Hall 2013).

In the stratified medicine approach the relative effect of treatment is allowed to vary across patients according to their treatment predictive factors. The relative increase in pregnancy chances for treatment in relation to no treatment has limited value since it does not tell us from what baseline chance (i.e. chance of pregnancy without treatment) the increase occurs. Stratified medicine considers the absolute rather than the relative increase in chance of pregnancy with treatment since the former provides the more relevant individualised prediction of successful treatment to guide decision-making.

Some thought needs to be given to identifying factors that predict differential treatment response. In fertility, the success of treatment, such as IVF, is heavily influenced by factors such as female age (van Loendersloot et al. 2010). As age is also a subfertility prognostic factor, increasing age may vary the additional effect of treatment over expectant management on chances of pregnancy. In other words, prognostic factors such as age, which affect the chance of spontaneous pregnancy and

success of IVF may also be treatment predictive factors which determine the relative effectiveness of treatment (Hingorani et al. 2013). Of interest is the difference in these two relative effects. Moreover, it is likely that an older woman whose chance of pregnancy with treatment is expected to be better than without, will require a more rapid resolution involving assisted reproduction, whilst a younger patient has sufficient time to undergo a series of less invasive (and cheaper) alternatives first. We know that as female age increases the ability of assisted reproduction technology to make up for all births lost by the natural decline of fertility decreases (Leridon 2004). Nevertheless, the absolute (and relative) benefit of treatment may be larger in older women than for younger women. There may also exist factors that are not necessarily prognostic that may predict the treatment response. For example, in women with different tubal factor subfertility problems those with hydrosalpinges had a poorer IVF pregnancy rate, which can be improved by salpingectomy (Johnson et al. 2011). Within such a cohort of women, subfertility prognosis would not be expected to vary between different tubal factor diagnoses, but type of tubal factor subfertility is clearly a treatment predictive factor.

Issues to consider for a stratified approach

A stratified model can be developed from: i. one data source that has compared treated versus untreated patient outcomes; or ii. two separate sources – one to model subfertility prognosis and one to predict outcome following treatment. We will discuss these in turn.

One data source, one model

This involves using one dataset, preferably from an RCT, comparing treatment with no treatment. One can then examine the effect of prognostic factors for subfertility (main effects in a statistical model) (Figure 1, Model 2a) together with treatment predictive factors (interaction terms in a statistical model) (Figure 1, Model 2b).

We could not find any examples of a published stratified medicine analysis for fertility using treatment predictive factors. However, a recent study attempted the secondary analysis of individual patient data from RCTs to determine whether a patient's prognostic profile, based on a score from the Hunault model, influenced the effectiveness of different fertility treatments (van den Boogaard et al. 2013). Investigating how the prognostic score from a model affects the treatment response, rather than the individual treatment predictive factors which made up the score, is called a risk-stratified analysis (Kent DM 2007).

Due to heterogeneity in the treatment protocols of the included trials in the Van den Boogaard study it was not possible to combine the individual patient data from each trial to conduct a meta-analysis. The modelling was performed in each trial separately. The study found no effect of prognostic profile on the effectiveness of different clinical strategies, including expectant management. This highlights the need for large RCTs with more heterogeneity in patient characteristics if they are to be used for secondary analyses involving modelling (Farooq et al. 2013). However, this is an expensive, challenging and lengthy process.

Although large RCTs are the preference for stratified medicine research, the use of observational data containing treated and non-treated women is an alternative. Such data usually contain a larger and more varied sample of patients than an RCT. An observational design requires high quality electronic healthcare data that can be record-linked in order to obtain an accurate history of the patient's journey (Hemingway *et al.,* 2013). However, observational data can suffer from serious selection bias issues, and whilst there are methods available that may be able to account for some of these, the results of any analyses should be interpreted with caution.

Two separate data sources, two models

In the absence of RCTs or observational databases containing both treated and non-treated women, a third approach is possible. This can use either previously published models – e.g. a prognostic model for spontaneous pregnancy, such as Hunault, and a model predicting treatment dependent pregnancy, such as the Nelson and Lawlor IVF model – or develop new models for each outcome using two separate data sources. The advantage of the former method is that most of the work has already been done and it is much less expensive than setting up a prospective, or even a retrospective, database from scratch. The difference in the absolute probability of success from both models would give the absolute benefit of treatment (Figure 1, Models 1a and 1b combined). However, a key problem with this method is the comparability of cohorts. The limitations of combining models developed from two different cohorts were highlighted in the recently updated National Institute of Clinical Excellence (NICE) clinical guideline on assessment and treatment for people with fertility problems (National Collaborating Centre for Women's and Children's Health 2013). A health economic analysis to compare the cost-effectiveness of different treatment strategies over a woman's reproductive life used the Hunault and the Nelson and Lawlor models to inform the cost-effectiveness model with probabilities of cumulative live birth in women following spontaneous pregnancy and IVF dependent pregnancy respectively. However, as the guideline acknowledges, there were major limitations associated with this approach. For example, the Hunault model was developed using a cohort of subfertile women, which excluded those who would not be expected to conceive naturally, meaning the severity of subfertility may not be as high as that in women referred for IVF (the cohort used for the Nelson and Lawlor models). Further, the maximum age of women used to develop the Hunault model was less than the maximum age included in the NICE cost-effectiveness model, which may result in an overestimate of the probability of spontaneous live birth in older aged women. However, if separate cohorts exist, which contain patients with very similar characteristics, who undergo either expectant management or treatment, then previous models can be adapted to fit such data or new models can be developed. If such

cohorts are available then this two-model approach would be equivalent to using the one model approach with statistical interaction terms between treatment and the treatment predictive factors.

When to treat? – A dynamic prediction approach

Another major aspect of clinical decision-making concerns the length of time couples should be advised to continue trying to conceive naturally before treatment should be offered. In order to do this we need a dynamic approach where we constantly assess the change in subfertility prognosis at different points in the future. One method is dynamic prediction modelling (van Houwelingen and Putter 2012). This involves fitting multiple time to event models from sequential equally spaced time points to predict natural pregnancy over, say, the following year (see Figure 2). This process enables one to determine the impact of delayed treatment on the predicted probability of pregnancy at different points in time. This is not the same as using, for example the Hunault model, to obtain the updated chances of pregnancy as time goes on by iteratively updating the same woman's prognostic factors for subfertility at baseline (i.e. when the cause of infertility is established). Rather, as time progresses the more fertile couples are excluded from the cohort due to pregnancy. Therefore after, for example, 6 months the cohort has reduced in size and is less fertile on average than the full sized cohort on which the model was originally based. Furthermore, since the original follow-up period has been extended by 6 months (i.e. follow-up now ends 18 months from baseline as opposed to 12 months) some of the women may have conceived during this period. Thus, different model estimates will be obtained.

Dynamic prediction could be used to advise those patients who are found to have a high chance of conceiving spontaneously at their first visit on when to return for treatment if their attempts are unsuccessful e.g. when their absolute chance of pregnancy dips below some pre-specified threshold.

It could also be used to make decisions regarding the immediate treatment for couples who have a low probability of pregnancy at their initial visit, which will decline further with each passing month.

Dynamic prediction should be used with the stratified medicine approach in order to estimate the change in the absolute benefit of treatment over time. In a couple with a good subfertility prognosis initially advised expectant management, this approach could be used to decide when in the future the absolute benefit of treatment is likely to trump their chance of spontaneous pregnancy such that the couple should be advised to return for treatment.

Dynamic prediction requires a cohort of patients with a sufficient length of follow-up to enable modelling at different time points. For this reason, existing observational datasets would be more suitable than an RCT. Finally, as for all clinical prediction modelling, the key steps involved in development and validation should be considered. The latter have been highlighted in the PROGRESS series (Steyerberg et al. 2013).

Practical recommendations

Given the complexities of the above approach to individualised-decision making in subfertility treatment, it is worth considering some practical guidelines for clinical practice and research. Firstly, the decision *whether* to treat a subfertile patient requires careful consideration of her background chance of spontaneous pregnancy and her predicted response to treatment. The former is influenced by prognostic factors and the latter by treatment predictive factors. Currently, in the Netherlands, an online prediction tool called 'Freya', based on the Hunault model, is used in clinical practice to make treatment decisions based on the probability of spontaneous ongoing pregnancy within the next 12 months (Hunault et al. 2004). However, clinicians should be aware that this model does not provide an estimate of response to treatment. Currently, the only way to do this is to use a combination of existing models from the literature, such as the Hunault model and the Nelson and

Lawlor model, which can be used to predict the chance of live birth following IVF. As mentioned earlier, this approach was used in a cost-effectiveness analysis of IVF relative to expectant management by NICE who acknowledge the shortcomings of this approach (National Collaborating Centre for Women's and Children's Health 2013).

Secondly, clinicians looking after couples with unexplained subfertility need to make a conscious decision as to *when* treatment should be offered. Depending on patient characteristics, such as female age, the live birth rate following one or more episodes of treatment will vary compared to what might be expected without treatment. Thus, it may be better to treat some women straight away after a diagnosis has been made, whilst in others a period of expectant management may lead to comparable or better live birth rates without the expense and invasiveness of active treatment. From the NICE analysis using the combined models, a 34 year old woman with two years of unexplained infertility is predicted to have a treatment independent live birth rate of 20% (National Collaborating Centre for Women's and Children's Health 2013) compared to 40% after one cycle of IVF. The same model predicts a live birth rate of 55% without treatment versus 70% following three complete cycles of IVF over the next 11 years, suggesting that it would seem advantageous to offer IVF treatment.

Finally, output from clinical predictive models need to be interpreted in the context of the individual circumstances of each couple. For fertility care to be genuinely patient centred, treatment decisions should involve couples themselves and accommodate their personal values and preferences (Dancet et al. 2011).

Conclusions

The current one-year definition of infertility should be used as a trigger for referral to the fertility clinic in order to initiate investigations and estimate prognosis – but not necessarily to begin treatment in all.

Current prognostic models in reproductive medicine are reasonably good at predicting the chances of pregnancy in either women who are treated or those who are not. As none of the existing models include both groups, predicting the marginal benefit of treatment versus no treatment is less accurate.

We advise the stratified medicine approach to identify those who actually benefit more from fertility treatment based on their prognostic and treatment predictive factors. Subsequently, the added benefit of treatment needs to be considered in context, for example in relation to the age of the woman. We also advise the dynamic prediction approach to estimate the patient's changing subfertility prognosis over time which could inform the decision about when to treat.

Further research needs to be undertaken to identify treatment predictive factors and to identify or create databases to allow these approaches to be explored. RCT data are preferred, but are the most challenging and expensive choice. In the interim, the most feasible option is to use output from a combination of previously published clinical prediction models, whilst acknowledging the specific clinical circumstances of each couple and their preferences.

Author's roles

DJM, SB and EWS proposed the concept. DJM drafted the paper and all named authors contributed content and commented on the draft.

Funding

This work was supported by a Chief Scientist Office Postdoctoral Training Fellowship in Health Services Research and Health of the Public Research (Ref PDF/12/06).

Conflict of interest

None of the authors declare any conflict of interest.

References

Bongaarts J. A Method for the Estimation of Fecundability. *Demography* 1975;**12**:645-660.

Collins JA. Associate editor's commentary: Mathematical modelling and clinical prediction. *Hum Reprod* 2005;**20**:2932-2934.

Coppus SFPJ, van der Veen F, Opmeer BC, Mol BWJ and Bossuyt PMM. Evaluating prediction models in reproductive medicine. *Hum Reprod* 2009;**24**:1774-1778.

Custers IM, Steures P, van der Steeg JW, van Dessel TJHM, Bernardus RE, Bourdrez P, Koks CAM, Riedijk WJ, Burggraaff JM, van der Veen F *et al*. External validation of a prediction model for an ongoing pregnancy after intrauterine insemination. *Fertil Steril* 2007;**88**:425-431.

Dancet EAF, Van Empel IWH, Rober P, Nelen WLDM, Kremer JAM, D'Hooghe TM. Patient-centred infertility care: a qualitative study to listen to the patient's voice. *Hum Reprod* 2011;**26**:827-833.

Evers J. Female subfertility. *The Lancet* 2002:**360**:151-159.

Farooq V, van Klaveren D, Steyerberg EW, Meliga E, Vergouwe Y, Chieffo A, Kappetein AP, Colombo A, Holmes Jr DR, Mack M *et al*. Anatomical and clinical characteristics to guide decision making

between coronary artery bypass surgery and percutaneous coronary intervention for individual patients: development and validation of SYNTAX score II. *The Lancet* 2013:**381**:639-650.

Gnoth C, Godehardt E, Frank-Herrmann P, Friol K, Tigges J and Freundl G. Definition and prevalence of subfertility and infertility. *Human Reproduction* 2005:**20**:1144-1147.

Greenhall E and Vessey M. The prevalence of subfertility: a review of the current confusion and a report of two new studies. *Fertil Steril* 1990:**54**:978-983.

Gurunath S, Pandian Z, Anderson RA and Bhattacharya S. Defining infertility - a systematic review of prevalence studies. *Human Reproduction Update* 2011:**17**:575-588.

Habbema JDF, Collins J, Leridon H, Evers JLH, Lunenfeld B and te Velde ER. Towards less confusing terminology in reproductive medicine: a proposal. *Human Reproduction* 2004:**19**:1497-1501.

Hall IP. Stratified medicine: drugs meet genetics. *Eur Respir Rev* 2013:**22**:53-57.

Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, Briggs A, Udumyan R, Moons KGM, Steyerberg EW *et al*. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ* 2013:**345**:e5595.

Hingorani AD, van der Windt DA, Riley RD, Abrams K, Moons KGM, Steyerberg EW, Schroter S, Sauerbrei W, Altman DG and Hemingway H. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ* 2013:**346**:e5793.

Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JLH and te Velde ER. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. *Human Reproduction* 2004:**19**:2019-2026.

391 Hunault CC, Laven JSE, van Rooij IAJ, Eijkemans MJC, te Velde ER and Habbema JDF. Prospective
392 validation of two models predicting pregnancy leading to live birth among untreated subfertile
393 couples. *Human Reproduction* 2005:**20**:1636-1641.

394 Italiano A. Prognostic or Predictive? It's Time to Get Back to Definitions!. *Journal of Clinical Oncology*
395 2011:**29**:4718-4718.

396 Johnson N, van Voorst S, Sowter MC, Strandell A and Mol BWJ. Tubal surgery before IVF. *Hum*
397 *Reprod Update* 2011:**17**:3-3.

398 Kent DM HR. Limitations of applying summary results of clinical trials to individual patients: The need
399 for risk stratification. *JAMA* 2007:**298**:1209-1212.

400 LaRosa JC, He J,Vupputuri S. Effect of statins on risk of coronary disease: A meta-analysis of
401 randomized controlled trials. *JAMA* 1999:**282**:2340-2346.

402 Leridon H. Can assisted reproduction technology compensate for the natural decline in fertility with
403 age? A model assessment. *Hum Reprod* 2004:**19**:1548-1553.

404 Leushuis E, van der Steeg JW, Steures P, Bossuyt PMM, Eijkemans MJC, van der Veen F, Mol BWJ and
405 Hompes PGA. Prediction models in reproductive medicine: a critical appraisal†. *Human Reproduction*
406 *Update* 2009:**15**:537-552.

407 National Collaborating Centre for Women's and Children's Health. Fertility: assessment and
408 treatment for people with fertility problems 2013:.

409 Nelson SM and Lawlor DA. Predicting Live Birth, Preterm Delivery, and Low Birth Weight in Infants
410 Born from In Vitro Fertilisation: A Prospective Study of 144,018 Treatment Cycles. *PLoS Med*
411 2011:**8**:e1000386.

412 Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, Malats N, Briggs A, Schroter
413 S, Altman DG *et al*. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLoS*
414 *Med* 2013:**10**:e1001380.

415 Smeenk JMJ, Stolwijk AM, Kremer JAM and Braat DDM. External validation of the Templeton model
416 for predicting success after IVF. *Human Reproduction* 2000:**15**:1065-1068.

417 Smeenk JMJ, Braat DDM, Stolwijk AM and Kremer JAM. Pregnancy is predictable: a large-scale
418 prospective external validation of the prediction of spontaneous pregnancy in subfertile couples.
419 *Human Reproduction* 2007:**22**:2344-2345.

420 Snick HK, Snick TS, Evers JL and Collins JA. The spontaneous pregnancy prognosis in untreated
421 subfertile couples: the Walcheren primary care study. *Human Reproduction* 1997:**12**:1582-1588.

422 Steures P, van der Steeg J, Hompes P, Habbema J, Eijkemans M, Broekmans F, Verhoeve H, Bossuyt
423 P, van der Veen F and Mol B. Intrauterine insemination with controlled ovarian hyperstimulation
424 versus expectant management for couples with unexplained subfertility and an intermediate
425 prognosis: a randomised clinical trial. *The Lancet* 2006:**368**:216-221.

426 Steures P, van der Steeg JW, Mol BWJ, Eijkemans MJC, van der Veen F, Habbema JDF, Hompes PGA,
427 Bossuyt PMM, Verhoeve HR, van Kasteren YM *et al*. Prediction of an ongoing pregnancy after
428 intrauterine insemination. *Fertil Steril* 2004:**82**:45-51.

429 Steyerberg E. Clinical prediction models: a practical approach to development, validation, and
430 updating. 1st edn, 2009. Springer, New York.

431 Steyerberg EW, Moons KGM, van dW, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H,
432 Altman DG and for the PG. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research.
433 *PLoS Med* 2013:**10**:e1001381.

te Velde ER, Eijkemans R and Habbema HD. Variation in couple fecundity and time to pregnancy, an essential concept in human reproduction. *Lancet* 2000:**355**:1928-1929.

te Velde ER, Nieboer D, Lintsen AM, Braat DDM, Eijkemans MJC, Habbema JDF and Vergouwe Y. Comparison of two UK models predicting IVF success; the effect of time trends on model performance. *Hum Reprod* 2014:**29**:57-64.

Templeton A, Morris JK and Parslow W. Factors that affect outcome of in-vitro fertilisation treatment. *The Lancet* 1996:**348**:1402-1406.

van den Boogaard NM, Bensdorp AJ, Oude Rengerink K, Barnhart K, Bhattacharya S, Custers IM, Coutifaris C, Goverde AJ, Guzick DS, Hughes EC *et al*. Prognostic profiles and the effectiveness of assisted conception: secondary analyses of individual patient data. *Human Reproduction Update* 2013:.

van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Hompes PGA, Broekmans FJ, van Dessel HJHM, Bossuyt PMM, van der Veen F and Mol BWJ. Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. *Human Reproduction* 2007:**22**:536-542.

van Houwelingen HC and Putter H. Dynamic prediction in clinical survival analysis. 1st edn, 2012. CRC Press, Boca Raton.

van Loendersloot LL, van Wely M, Limpens J, Bossuyt PMM, Repping S and van der Veen F. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. *Human Reproduction Update* 2010:**16**:577-589.

Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, Sullivan E and van der Poel S. The International Committee for Monitoring Assisted Reproductive Technology (ICMART)

and the World Health Organization (WHO) Revised Glossary on ART Terminology, 2009. *Hum Reprod* 2009:**24**:2683-2687.

475 **Table 1**. Prognostic factors used to predict spontaneous pregnancy (taken from Leushuis et al, 2009).

Couple factors
Duration of subfertility (year)
Secondary subfertility
Female factors
Female age (years)
Referral status (tertiary care)
Ovulation disorder
Pelvic surgery
Tubal defect
Endometriosis
Ovulation or cervical disorder
Uterine abnormality (UA)
UA and ovulation or cervical disorder
Male factors
Male age (year)
Sperm motility (%)
Degree of motility (good)
Sperm morphology (%)
Sperm concentration (x 10^6)
Abnormal post coital test (PCT)
World Health Organisation (WHO) semen defect
Hypo-osmotic test (HOS) test (%)
Urethritis in history
Fertility problem in male's family

476

Figure 1 Diagram to explain absolute and relative benefit of treatment (Tx) in the stratified medicine approach for individualised predictions of a pregnancy outcome, such as live birth, in a subfertile population

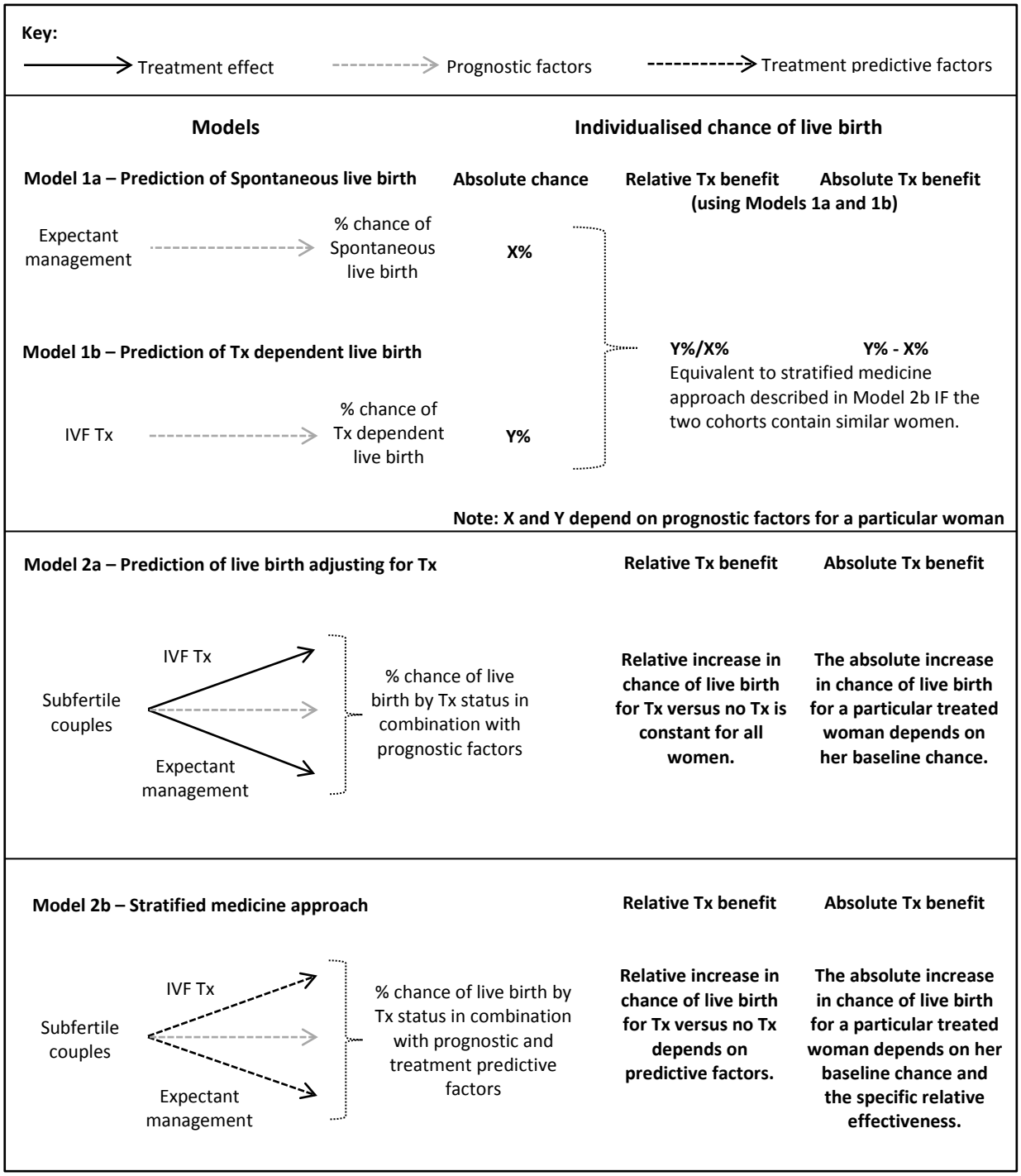
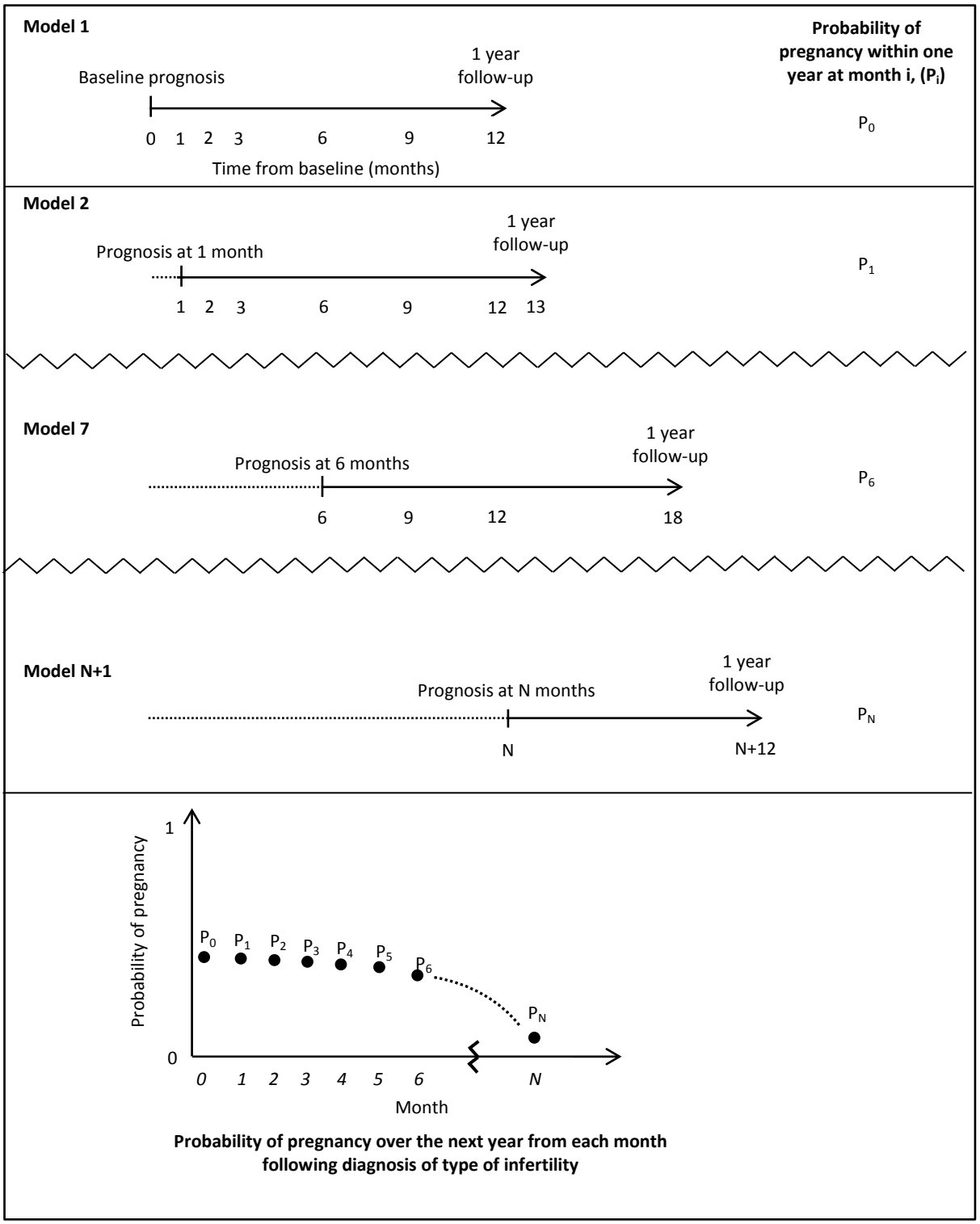


Figure 2 Dynamic prediction for pregnancy prognosis



Model 1: A time to event model predicts the probability of pregnancy (P_0) within one year at the point where the type of infertility is established (baseline).

Model 2: A second time to event model predicts the probability of pregnancy (P_1) within one year from 1 month after baseline. All women who were pregnant in the first month (dotted line) are excluded.
This is repeated from every month thereafter, until month N.

Model *N*+1: An *(N+1)*th time to event model predicts the probability of pregnancy (P_N) within one year from *N* months after baseline. All women who were pregnant up to month *N* (dotted line) are excluded.