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

D.J. McLernon1,*, E.R. te Velde2, E.W. Steyerberg2, B.W.J. Mol3, and S. Bhattacharya4

1Medical Statistics Team, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, UK
2Department of Public Health, Erasmus MC – University Medical Centre Rotterdam, Rotterdam, The Netherlands 3The Robinson Institute, School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, Australia 4Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, UK

*Correspondence address. Tel: +44-1224437152; Fax: +44-1224437285; E-mail: d.mclernon@abdn.ac.uk

Submitted on March 14, 2014; resubmitted on May 8, 2014; accepted on June 5, 2014

ABSTRACT: Infertility is defined as failure to conceive after 1 year of unprotected intercourse. This dichotomization 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 individualized 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 personalized 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.

Key words: female infertility / pregnancy / statistics / epidemiology

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 Organization (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’ (Habbema et al., 2004; Gnoth et al., 2005; Gurunath et al., 2011). 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 azoosperma. 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 1 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 recognizes 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 look at the probability 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 with 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: first, 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 I 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 8 were externally validated, 3 of which showed adequate performance (Templeton et al., 1996; Smeenk et al., 2000; Hunault et al., 2004; Steures et al., 2004, 2006; Custers et al., 2007; 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, synthesized from 3 previous models based on 3 prospective databases of subfertile women attending a Dutch University hospital, a Dutch general hospital and 11 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.

<table>
<thead>
<tr>
<th>Table I Prognostic factors used to predict spontaneous pregnancy (taken from Leushuis et al., 2009, by permission of Oxford University Press).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Couple factors</strong></td>
</tr>
<tr>
<td>Duration of subfertility (year)</td>
</tr>
<tr>
<td>Secondary subfertility</td>
</tr>
<tr>
<td><strong>Female factors</strong></td>
</tr>
<tr>
<td>Female age (years)</td>
</tr>
<tr>
<td>Referral status (tertiary care)</td>
</tr>
<tr>
<td>Ovulation disorder</td>
</tr>
<tr>
<td>Pelvic surgery</td>
</tr>
<tr>
<td>Tubal defect</td>
</tr>
<tr>
<td>Endometriosis</td>
</tr>
<tr>
<td>Ovulation or cervical disorder</td>
</tr>
<tr>
<td>Uterine abnormality (UA)</td>
</tr>
<tr>
<td>UA and ovulation or cervical disorder</td>
</tr>
<tr>
<td><strong>Male factors</strong></td>
</tr>
<tr>
<td>Male age (year)</td>
</tr>
<tr>
<td>Sperm motility (%)</td>
</tr>
<tr>
<td>Degree of motility (good)</td>
</tr>
<tr>
<td>Sperm morphology (%)</td>
</tr>
<tr>
<td>Sperm concentration (x 10⁶)</td>
</tr>
<tr>
<td>Abnormal post coital test</td>
</tr>
<tr>
<td>WHO semen defect</td>
</tr>
<tr>
<td>Hypo-osmotic test test (%)</td>
</tr>
<tr>
<td>Urethritis in history</td>
</tr>
<tr>
<td><strong>Fertility problem in male’s family</strong></td>
</tr>
</tbody>
</table>
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.’s (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 (Smeenk et al., 2007; Coppus et al., 2009). 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 Fertilization and Embryological Authority 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. 1 year, for a given individual (see Fig. 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 (Fig. 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 with 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 optimize 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 individualized 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 Loendersloo 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

---

**Table 1**

<table>
<thead>
<tr>
<th>Models</th>
<th>Individualised chance of live birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1a – Prediction of Spontaneous live birth</td>
<td>Absolute chance</td>
</tr>
<tr>
<td>Expectant management</td>
<td>% chance of Spontaneous live birth</td>
</tr>
</tbody>
</table>

**Model 1b – Prediction of Tx dependent live birth**
| | % chance of Tx dependent live birth | Y% |

<table>
<thead>
<tr>
<th>Model 2a – Prediction of live birth adjusting for Tx</th>
<th>Relative Tx benefit</th>
<th>Absolute Tx benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subfertile couples</td>
<td>% chance of live birth by Tx status in combination with prognostic factors</td>
<td>Relative increase in chance of live birth for Tx versus no Tx is constant for all women.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 2b – Stratified medicine approach</th>
<th>Relative Tx benefit</th>
<th>Absolute Tx benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subfertile couples</td>
<td>% chance of live birth by Tx status in combination with prognostic and treatment predictive factors</td>
<td>Relative increase in chance of live birth for Tx versus no Tx depends on predictive factors.</td>
</tr>
</tbody>
</table>

---

**Diagram**

![Diagram to explain absolute and relative benefit of treatment (Tx) in the stratified medicine approach for individualized predictions of a pregnancy outcome, such as live birth, in a subfertile population.](https://academic.oup.com/humrep/article-abstract/29/9/1851/2427952/1854)

---

**Figure 1**

Diagram to explain absolute and relative benefit of treatment (Tx) in the stratified medicine approach for individualized predictions of a pregnancy outcome, such as live birth, in a subfertile population.
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) (Fig. 1, Model 2a) together with treatment predictive factors (interaction terms in a statistical model) (Fig. 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, 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 (Fig. 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 Fig. 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.
Figure 2. Dynamic prediction for pregnancy prognosis. Model 1: a time to event model predicts the probability of pregnancy ($P_0$) within 1 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 1 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 1 year from $N$ months after baseline. All women who were pregnant up to month $N$ (dotted line) are excluded.
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 individualized decision-making in subfertility treatment, it is worth considering some practical guidelines for clinical practice and research. First, 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 with 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 2 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 with 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 1-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 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.

**Authors’ roles**

D.J.M., S.B. and E.W.S. proposed the concept. D.J.M. 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). The views expressed in this paper represent the views of the authors and not necessarily the views of the funding body.

**Conflict of interest**

None of the authors declare any conflict of interest.

**References**


Italiano A. Progностico or Predictive? It’s time to get back to definitions! J Clin Oncol 2011;29:4718—4718.
Kent DM. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. JAMA 2007;298:1209—1212.