

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/114437/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

McLernon, D. J., te Velde, E. R., Steyerberg, E. W., Mol, B. W. J. and Bhattacharya, S. 2014. Clinical prediction models to inform individualized decision-making in subfertile couples: a stratified medicine approach. *Human Reproduction* 29 (9) , pp. 1851-1858. 10.1093/humrep/deu173

Publishers page: <http://dx.doi.org/10.1093/humrep/deu173>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



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

3 **DJ McLernon<sup>1</sup>, ER te Velde<sup>2</sup>, EW Steyerberg<sup>2</sup>, BWJ Mol<sup>3</sup>, S Bhattacharya<sup>1</sup>**

4 <sup>1</sup>Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK; <sup>2</sup>Department of Public  
5 Health, Erasmus MC - University Medical Centre Rotterdam, Rotterdam, The Netherlands;

6 <sup>3</sup>Department of Obstetrics and Gynaecology, Academic Medical Centre, Amsterdam, The  
7 Netherlands.

8

9 Corresponding author:

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

12 Tel: +441224437152, Fax: +441224437285

13 Email: d.mclernon@abdn.ac.uk

14 Running title: Individualized decision making in subfertility

15 Word Count: 4139

16

**17 Abstract**

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

## 38 **Introduction**

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

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

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

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

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

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

83

#### 84 **Existing clinical prediction models in subfertility**

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

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

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

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

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

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

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

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

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

146

#### 147 **Absolute versus relative risk**

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

159

#### 160 **To treat or not to treat? - A stratified medicine approach**



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

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

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

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

200

#### 201 **Issues to consider for a stratified approach**

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

#### 205 ***One data source, one model***

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

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

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

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

231

232 ***Two separate data sources, two models***

233 In the absence of RCTs or observational databases containing both treated and non-treated women,  
234 a third approach is possible. This can use either previously published models – e.g. a prognostic  
235 model for spontaneous pregnancy, such as Hunault, and a model predicting treatment dependent  
236 pregnancy, such as the Nelson and Lawlor IVF model – or develop new models for each outcome  
237 using two separate data sources. The advantage of the former method is that most of the work has  
238 already been done and it is much less expensive than setting up a prospective, or even a  
239 retrospective, database from scratch. The difference in the absolute probability of success from both  
240 models would give the absolute benefit of treatment (Figure 1, Models 1a and 1b combined).  
241 However, a key problem with this method is the comparability of cohorts. The limitations of  
242 combining models developed from two different cohorts were highlighted in the recently updated  
243 National Institute of Clinical Excellence (NICE) clinical guideline on assessment and treatment for  
244 people with fertility problems (National Collaborating Centre for Women's and Children's Health  
245 2013). A health economic analysis to compare the cost-effectiveness of different treatment  
246 strategies over a woman's reproductive life used the Hunault and the Nelson and Lawlor models to  
247 inform the cost-effectiveness model with probabilities of cumulative live birth in women following  
248 spontaneous pregnancy and IVF dependent pregnancy respectively. However, as the guideline  
249 acknowledges, there were major limitations associated with this approach. For example, the Hunault  
250 model was developed using a cohort of subfertile women, which excluded those who would not be  
251 expected to conceive naturally, meaning the severity of subfertility may not be as high as that in  
252 women referred for IVF (the cohort used for the Nelson and Lawlor models). Further, the maximum  
253 age of women used to develop the Hunault model was less than the maximum age included in the  
254 NICE cost-effectiveness model, which may result in an overestimate of the probability of  
255 spontaneous live birth in older aged women. However, if separate cohorts exist, which contain  
256 patients with very similar characteristics, who undergo either expectant management or treatment,  
257 then previous models can be adapted to fit such data or new models can be developed. If such

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

260

### 261 **When to treat? – A dynamic prediction approach**

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

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

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

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

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

293

#### 294 **Practical recommendations**

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

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

309

310 Secondly, clinicians looking after couples with unexplained subfertility need to make a conscious  
311 decision as to *when* treatment should be offered. Depending on patient characteristics, such as  
312 female age, the live birth rate following one or more episodes of treatment will vary compared to  
313 what might be expected without treatment. Thus, it may be better to treat some women straight  
314 away after a diagnosis has been made, whilst in others a period of expectant management may lead  
315 to comparable or better live birth rates without the expense and invasiveness of active treatment.

316 From the NICE analysis using the combined models, a 34 year old woman with two years of  
317 unexplained infertility is predicted to have a treatment independent live birth rate of 20% (National  
318 Collaborating Centre for Women's and Children's Health 2013) compared to 40% after one cycle of  
319 IVF. The same model predicts a live birth rate of 55% without treatment versus 70% following three  
320 complete cycles of IVF over the next 11 years, suggesting that it would seem advantageous to offer  
321 IVF treatment.

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

326

327 **Conclusions**

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

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

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

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

345

#### 346 **Author's roles**

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

349

#### 350 **Funding**



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

353

#### 354 **Conflict of interest**

355 None of the authors declare any conflict of interest.

356

#### 357 **References**

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

359 Collins JA. Associate editor's commentary: Mathematical modelling and clinical prediction. *Hum*

360 *Reprod* 2005;**20**:2932-2934.

361 Coppus SFPJ, van der Veen F, Opmeer BC, Mol BWJ and Bossuyt PMM. Evaluating prediction models

362 in reproductive medicine. *Hum Reprod* 2009;**24**:1774-1778.

363 Custers IM, Steures P, van der Steeg JW, van Dessel TJHM, Bernardus RE, Bourdrez P, Koks CAM,

364 Riedijk WJ, Burggraaff JM, van der Veen F *et al*. External validation of a prediction model for an

365 ongoing pregnancy after intrauterine insemination. *Fertil Steril* 2007;**88**:425-431.

366 Dancet EAF, Van Empel IWH, Rober P, Nelen WLDM, Kremer JAM, D'Hooghe TM. Patient-centred

367 infertility care: a qualitative study to listen to the patient's voice. *Hum Reprod* 2011;**26**:827-833.

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

369 Farooq V, van Klaveren D, Steyerberg EW, Meliga E, Vergouwe Y, Chieffo A, Kappetein AP, Colombo

370 A, Holmes Jr DR, Mack M *et al*. Anatomical and clinical characteristics to guide decision making

- 371 between coronary artery bypass surgery and percutaneous coronary intervention for individual  
372 patients: development and validation of SYNTAX score II. *The Lancet* 2013;**381**:639-650.
- 373 Gnoth C, Godehardt E, Frank-Herrmann P, Friol K, Tigges J and Freundl G. Definition and prevalence  
374 of subfertility and infertility. *Human Reproduction* 2005;**20**:1144-1147.
- 375 Greenhall E and Vessey M. The prevalence of subfertility: a review of the current confusion and a  
376 report of two new studies. *Fertil Steril* 1990;**54**:978-983.
- 377 Gurunath S, Pandian Z, Anderson RA and Bhattacharya S. Defining infertility - a systematic review of  
378 prevalence studies. *Human Reproduction Update* 2011;**17**:575-588.
- 379 Habbema JDF, Collins J, Leridon H, Evers JLH, Lunenfeld B and te Velde ER. Towards less confusing  
380 terminology in reproductive medicine: a proposal. *Human Reproduction* 2004;**19**:1497-1501.
- 381 Hall IP. Stratified medicine: drugs meet genetics. *Eur Respir Rev* 2013;**22**:53-57.
- 382 Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, Briggs A, Udumyan R, Moons KGM,  
383 Steyerberg EW *et al.* Prognosis research strategy (PROGRESS) 1: A framework for researching clinical  
384 outcomes. *BMJ* 2013;**345**:e5595.
- 385 Hingorani AD, van der Windt DA, Riley RD, Abrams K, Moons KGM, Steyerberg EW, Schroter S,  
386 Sauerbrei W, Altman DG and Hemingway H. Prognosis research strategy (PROGRESS) 4: Stratified  
387 medicine research. *BMJ* 2013;**346**:e5793.
- 388 Hunault CC, Habbema JDF, Eijkemans MJC, Collins JA, Evers JLH and te Velde ER. Two new prediction  
389 rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the  
390 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.

- 434 te Velde ER, Eijkemans R and Habbema HD. Variation in couple fecundity and time to pregnancy, an  
435 essential concept in human reproduction. *Lancet* 2000;**355**:1928-1929.
- 436 te Velde ER, Nieboer D, Lintsen AM, Braat DDM, Eijkemans MJC, Habbema JDF and Vergouwe Y.  
437 Comparison of two UK models predicting IVF success; the effect of time trends on model  
438 performance. *Hum Reprod* 2014;**29**:57-64.
- 439 Templeton A, Morris JK and Parslow W. Factors that affect outcome of in-vitro fertilisation  
440 treatment. *The Lancet* 1996;**348**:1402-1406.
- 441 van den Boogaard NM, Bendsorp AJ, Oude Rengerink K, Barnhart K, Bhattacharya S, Custers IM,  
442 Coutifaris C, Goverde AJ, Guzick DS, Hughes EC *et al*. Prognostic profiles and the effectiveness of  
443 assisted conception: secondary analyses of individual patient data. *Human Reproduction Update*  
444 2013:.
- 445 van der Steeg JW, Steures P, Eijkemans MJC, Habbema JDF, Hompes PGA, Broekmans FJ, van Dessel  
446 HJHM, Bossuyt PMM, van der Veen F and Mol BWJ. Pregnancy is predictable: a large-scale  
447 prospective external validation of the prediction of spontaneous pregnancy in subfertile couples.  
448 *Human Reproduction* 2007;**22**:536-542.
- 449 van Houwelingen HC and Putter H. Dynamic prediction in clinical survival analysis. 1st edn, 2012. CRC  
450 Press, Boca Raton.
- 451 van Loendersloot LL, van Wely M, Limpens J, Bossuyt PMM, Repping S and van der Veen F. Predictive  
452 factors in in vitro fertilization (IVF): a systematic review and meta-analysis. *Human Reproduction*  
453 *Update* 2010;**16**:577-589.
- 454 Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, Sullivan E and van  
455 der Poel S. The International Committee for Monitoring Assisted Reproductive Technology (ICMART)

456 and the World Health Organization (WHO) Revised Glossary on ART Terminology, 2009. *Hum Reprod*  
457 2009;**24**:2683-2687.

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

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

<b>Couple factors</b>
Duration of subfertility (year)
Secondary subfertility
<b>Female factors</b>
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
<b>Male factors</b>
Male age (year)
Sperm motility (%)
Degree of motility (good)
Sperm morphology (%)
Sperm concentration ( $\times 10^6$ )
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

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

480

481

482

483

484

485

486

487

488

489

490

491

492

493

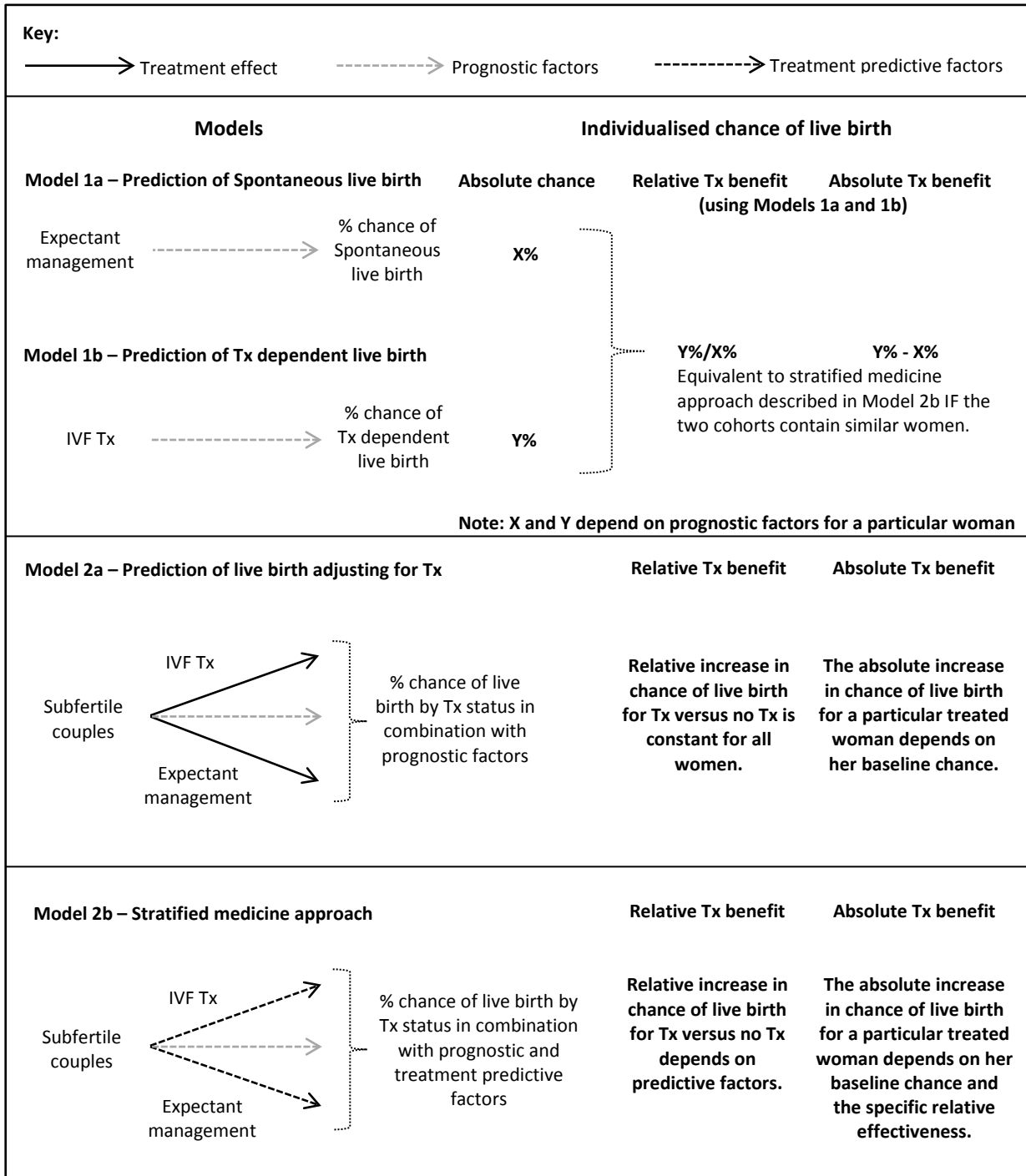
494

495

496

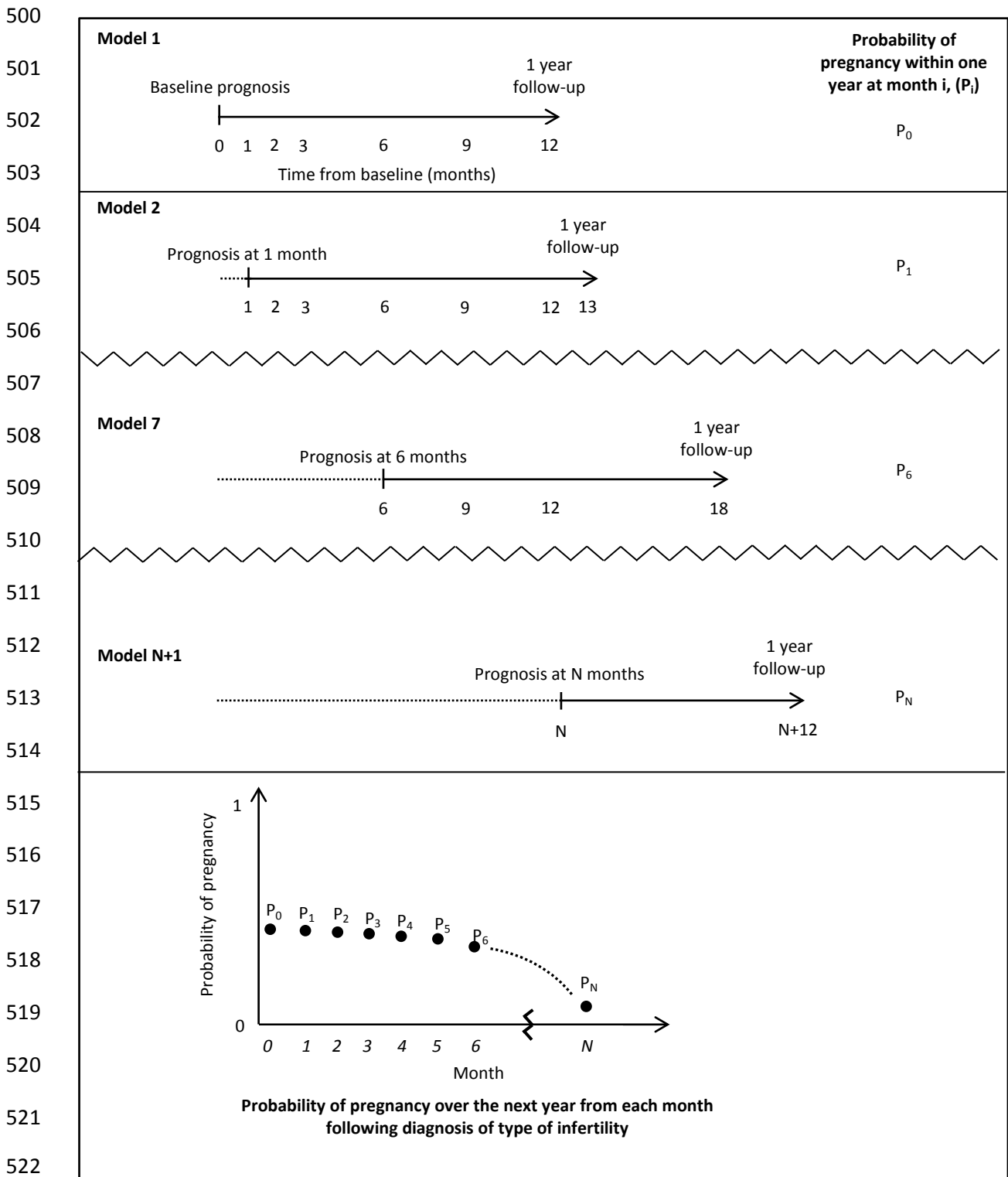
497

498





499 **Figure 2** Dynamic prediction for pregnancy prognosis



524 Model 1: A time to event model predicts the probability of pregnancy ( $P_0$ ) within one year at the point where the type of  
 525 infertility is established (baseline).  
 526 Model 2: A second time to event model predicts the probability of pregnancy ( $P_1$ ) within one year from 1 month after  
 527 baseline. All women who were pregnant in the first month (dotted line) are excluded.  
 528 This is repeated from every month thereafter, until month  $N$ .  
 529 Model  $N+1$ : An  $(N+1)$ th time to event model predicts the probability of pregnancy ( $P_N$ ) within one year from  $N$  months  
 530 after baseline. All women who were pregnant up to month  $N$  (dotted line) are excluded.