External validation of a dynamic prediction model for repeated predictions of natural conception over time

Journal:	*Human Reproduction*
Manuscript ID	HUMREP-18-0188.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	25-Sep-2018
Complete List of Authors:	van Eekelen, Rik; Amsterdam UMC, Academic Medical Center, Centre for Reproductive Medicine; University Medical Center Utrecht, Department of Biostatistics and Research Support, Julius Center McLernon, David; University of Aberdeen, Medical Statistics Team, Institute of Applied Health Sciences van Wely, Madelon; Amsterdam UMC, Amsterdam Medical Center, Centre for Reproductive Medicine Eijkemans, Marinus; University Medical Center Utrecht, Department of Biostatistics and Research Support, Julius Center Bhattacharya, Siladitya; University of Aberdeen, Institute of Applied Health Sciences van der Veen, Fulco; Amsterdam UMC, Academic Medical Center, Centre for Reproductive Medicine van Geloven, Nan; Leiden University Medical Center, Medical Statistics, Department of Biomedical Data Sciences
Keywords:	Natural conception, Expectant management, Prognosis, Prediction model, Dynamic prediction
Subject Section:	Infertility

SCHOLARONE™
Manuscripts

1. **External validation of a dynamic prediction model for**

2. **repeated predictions of natural conception over time**

3. **van Eekelen R[1*,2], McLernon DJ[3], van Wely M[1], Eijkemans MJ[2], Bhattacharya S[4], van**

4. **der Veen F[1], van Geloven N[5]**

5.

6. [1] Centre for Reproductive Medicine, Amsterdam UMC, Academic Medical Centre,

7. Meibergdreef 9 1105 AZ Amsterdam, the Netherlands

8. [2] Department of Biostatistics and Research Support, Julius Centre, University Medical Centre

9. Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands

10. [3] Medical Statistics Team, Institute of Applied Health Sciences, University of Aberdeen, AB24

11. 3FX Aberdeen, United Kingdom

12. [4] Institute of Applied Health Sciences, University of Aberdeen, AB24 3FX Aberdeen, United

13. Kingdom

14. [5] Medical Statistics, Department of Biomedical Sciences, Leiden University Medical Centre,

15. Einthovenweg 20, 2333 ZC Leiden, the Netherlands

16.

17.   •   Correspondence address. Email: r.vaneekelen@amc.uva.nl

18   **Running title:** External validation of a dynamic prediction model

19

20   **Abstract**

21

22   **STUDY QUESTION:** How well does a previously developed dynamic prediction model

23   perform in an external, geographical validation in terms of predicting the chances of natural

24   conception at various points in time?

25

26   **SUMMARY ANSWER:** The dynamic prediction model performs well in an external validation

27   on a Scottish cohort.

28

29   **WHAT IS KNOWN ALREADY:** Prediction models provide information that can aid evidence-

30   based management of unexplained subfertile couples. We developed a dynamic prediction

31   model for natural conception (van Eekelen model) that is able to update predictions of natural

32   conception when couples return to their clinician after a period of unsuccessful expectant

33   management. It is not known how well this model performs in an external population.

34

35   **STUDY DESIGN, SIZE, DURATION:** A record-linked registry study including the long-term

36   follow up of all couples who were considered unexplained subfertile following a fertility work

37   up at a Scottish fertility clinic between 1998 and 2011. Couples with anovulation, uni/bilateral

38   tubal occlusion, mild/severe endometriosis or impaired semen quality according to World

39   Health Organization criteria were excluded.

40

41   **PARTICIPANTS/MATERIALS, SETTING, METHODS**: The endpoint was time to natural

42   conception, leading to an ongoing pregnancy (defined as reaching a gestational age of at

43   least 12 weeks). Follow up was censored at the start of treatment, at the change of partner or

44   at the end of study (31st of March, 2012). The performance of the van Eekelen model was

45   evaluated in terms of calibration and discrimination at various points in time. Additionally, we

46   assessed the clinical utility of the model in terms of the range of the calculated predictions.

47

48   **MAIN RESULTS AND THE ROLE OF CHANCE:** Of a total of 1203 couples with a median

49   follow up of 1 year and 3 months after the fertility workup, 398 (33%) couples conceived

50   naturally leading to an ongoing pregnancy. Using the dynamic prediction model, the mean

51   probability of natural conception over the course of the first year after the fertility workup was

52   estimated at 25% (observed: 23%). After 0.5 year, 1 year and 1.5 years of expectant

53   management after completion of the fertility workup, the average probability of conceiving

54   naturally over the next year was estimated at 18% (observed: 15%), 14% (observed: 14%)

55   and 12% (observed: 12%).

56   Calibration plots showed good agreement between predicted chances and the observed

57   fraction of ongoing pregnancy within risk groups. Discrimination was moderate with c

58   statistics similar to those in the internal validation, ranging from 0.60 to 0.64. The range of

59   predicted chances was sufficiently wide to distinguish between couples having a good and

60   poor prognosis with a minimum of zero at all times and a maximum of 55% over the first year

61   after the workup, which decreased to maxima of 43% after 0.5 years, 34% after 1 year and

62   29% after1.5 years after the fertility workup.

63

64   **LIMITATIONS, REASONS FOR CAUTION:** The model slightly overestimated the chances of

65   conception by approximately 2 to 3 percentage points on group level in the first year post

66   fertility workup and after 0.5 years of expectant management, respectively. This is likely

67   attributable to the fact that the exact dates of completion of the fertility workup for couples

68   were missing and had to be estimated.

69

70   **WIDER IMPLICATIONS OF THE FINDINGS:** The van Eekelen model is a valid and robust

71   tool that is ready to use in clinical practice to counsel couples with unexplained subfertility on

72   their individualised chances of natural conception at various points in time, notably when

73   couples return to the clinic after a period of unsuccessful expectant management.

74

78

82   **Introduction**

83   Approximately 10% of all couples who wish to have a child do not conceive within the first

84   year of trying (Gnoth *et al.*, 2003; Wang *et al.*, 2003). For approximately half of these

85   couples, no clear barrier for conception can be found during the workup and these couples

86   are considered unexplained subfertile (Aboulghar *et al.*, 2009; Brandes *et al.*, 2010). It is

87   unclear whether these couples should start with ART; firstly, since observational studies

88   report that 18% to 38% of unexplained subfertile couples will conceive naturally in the year

89   after the fertility workup (Hunault *et al.*, 2004; van der Steeg *et al.*, 2007; van Eekelen *et al.*,

90   2017a) and secondly, since there remains uncertainty regarding the effectiveness of ART for

91   unexplained subfertile couples (Pandian *et al.*, 2015; Tjon-Kon-Fat *et al.*, 2016; Veltman-

92   Verhulst *et al.*, 2016; van Eekelen *et al.*, 2017b).

93   In the absence of clear evidence on the management of unexplained subfertile couples and

94   when to offer ART, an enticing option is to calculate chances of natural conception and to

95   base counselling on this estimated prognosis (van Eekelen *et al.*, 2017b). Fundamental to

96   this approach is to identify couples that are expected to benefit from treatment and those

97   who are not. In clinical practice, this would imply that couples with a good prognosis to

98   conceive naturally are advised to continue to try and become pregnant by sexual intercourse,

99   while couples with an unfavourable prognosis are advised to start ART. Several prediction

100   models for natural conception have been published of which the model by Hunault *et al.*, that

101   calculates a prognosis of conception leading to live birth over the first year after completion

102   of the fertility workup, has been externally validated and subsequently implemented in the

103   national guidelines and clinical practice in the Netherlands (Hunault *et al.*, 2004; van der

104   Steeg *et al.*, 2007; Leushuis *et al.*, 2009; NVOG, 2010). A practical drawback of the Hunault

105   model is that it cannot give a prediction at later time points when couples who continued

106   expectant management after the fertility workup but did not conceive, return to the clinic. This

107   is because applying the Hunault model at later time points leads to overestimation due to the

108   selection of less fertile couples over time that is not incorporated in the Hunault model (van

109   Eekelen *et al.*, 2017b).

110   Van Eekelen *et al*. recently developed a dynamic prediction model that accommodates the

111   need for repeated predictions (van Eekelen *et al.*, 2017a). This model comprises the clinical

112   factors female age, duration of subfertility (both at completion of the fertility workup),

113   percentage of progressively motile sperm, primary or secondary subfertility and being

114   referred to the fertility clinic by a general practitioner or a specialist. In addition to these

115   factors, the model uses as input the number of menstrual cycles that have passed since

116   completion of the fertility workup, with zero cycles denoting the prediction is made

117   immediately after the workup. The output is the predicted probability to conceive naturally in

118   the following cycle, leading to ongoing pregnancy, which can be extended to predict over any

119   given number of cycles with a maximum of 2.5 years after the workup (approximately 28-34

120   cycles). When couples return after a period of expectant management, the number of cycles

121   that have passed since the workup can be changed to update the predicted probability over

122   subsequent cycles.

123   The model developed by van Eekelen *et al*. showed promising results in the internal

124   validation, but this in itself is insufficient to advise clinical implementation since models tend

125   to perform better in the cohort they were developed on than in another cohort in which the

126   model may be applied (Steyerberg, 2009).

127   The aim of this study was to externally validate the van Eekelen model on a large cohort that

128   followed couples for natural conception after registration in the fertility clinic of the Grampian

129   region of Scotland, UK. This is the largest contemporary cohort following couples for natural

130   conception, aside from the Dutch cohort on which the dynamic model was developed.

131

132

133   **Materials and Methods**

134   We included couples diagnosed with unexplained subfertility residing in the Grampian region

135   of Scotland who registered with the Aberdeen Fertility Centre (AFC) from 1998 to 2011

136   (Pandey *et al.*, 2014). Only patients from the Grampian region visiting the AFC were selected

137   because there is no other fertility clinic in the region and it was considered important to have

138   a complete overview of a couple's trajectory after the fertility workup, which includes

139   treatment information. We combined the AFC registration database with three other data

140   sources using record-linkage to get the complete follow up for couples from the registration

141   at the AFC until ongoing pregnancy, treatment or end of study, which was the 31st of March,

142   2012.

143   The AFC database comprises patient characteristics and diagnostic information. Data entry

144   in the AFC database is validated and checked by regular case note audits. First, we record-

145   linked couples registered in the AFC database to the centre's Assisted Reproduction Unit

146   database which contained dates when treatment was started.

147   Second, we identified natural conceptions leading to an ongoing pregnancy by record-linkage

148   of the AFC database with the Aberdeen Maternity and Neonatal Databank, which contained

149   gestational age, outcome and delivery date of (early) pregnancies for all women residing in

150   Aberdeen City District. Third, we performed record-linkage with the national Scottish

151   Morbidity Records Maternity database for identifying gestational age, outcome and delivery

152   date of (early) pregnancies for women who delivered elsewhere in Scotland.

153   The Data Management Team of the University of Aberdeen created a new pseudonomised

154   identifier for all women by using the Community Health Index identifier. This new study-

155   specific identifier cannot be used to trace back to individuals and was then used by author

156   DJM to record-link the databases within the Grampian Data Safe Haven environment. This

157   process was carried out according to the Standard Operating Procedures of the Data

158   Management Team, University of Aberdeen. The resulting linked dataset was thus a

159   combination of these four data sources.

160   Ethical approval was provided by the North of Scotland Research Ethics Committee

161   (reference: 12/NS/0120). Access to the Aberdeen Fertility Clinic and the Assisted

162   Reproduction Unit databases was approved by the Aberdeen Fertility Databases Steering

163   Committee. Access to the Aberdeen Maternity and Neonatal Databank was approved by the

164   Aberdeen Maternity and Neonatal Database Steering Committee. Access to the Scottish

165   Morbidity Records Maternity database was approved by the Privacy Advisory Committee of

166   Information Services Division Scotland.

167   We defined unexplained subfertility as couples who tried to conceive for more than 50 weeks

168   before the fertility workup was completed and who had no obvious barriers to conception in

169   terms of uni- or bilateral tubal occlusion, anovulation, mild- or severe endometriosis

170   according to the revised American Society for Reproductive Medicine (ASRM) score (ASRM,

171   1997) or impaired semen quality according to World Health Organization (WHO) criteria

172   (WHO, 1999; WHO, 2010). We used the gestational age at birth or early pregnancy outcome

173   to derive the date of conception and included only pregnancies in the analysis that occurred

174   after registration of the couple at the clinic and that were ongoing, defined as reaching a

175   gestational age of at least 12 weeks. Time to conception was censored at the date of start of

176   IUI, start of IVF, when the woman returned to the fertility centre with a different male partner

177   or at the end of study.

178

179   *Missing data*

180   The date of completion of the fertility workup was not reported in the AFC database. The van

181   Eekelen model uses this date as the starting point of follow up, i.e. the time point from which

182   onwards the model can be used to estimate a prognosis. The date of registration and the

183   diagnosis category were available in the database. Judging from local protocols, we

184   assumed there were 3 months in between registration and completion of the fertility workup

185   for all couples. In a sensitivity analysis, we repeated the validation study assuming 1.5

186   months or 4.5 months between registration and completion of the fertility workup for all

187   couples.

188   Menstrual cycle length is used to determine the number of elapsed menstrual cycles since

189   the fertility workup when updating predictions using the dynamic prediction model. Cycle

190   length was not recorded in the AFC database and we therefore assumed an average cycle

191   length of 28 days for all women.

192   Data on outcomes or at least one prognostic factor were missing for approximately 4% of

193   couples; 0.5% on pregnancy or follow up, 0.5% on female age, 2.3% on duration of

194   subfertility, 0.5% on primary or secondary subfertility, 1.9% on the percentage of progressive

195   motile sperm and 0.5% on referral status. We had no reason to believe that couples with

196   missing data differed systematically from couples with complete data and we analysed

197   couples for which data was complete.

198

199   *Analysis*

200   We calculated the predicted probabilities of natural conception over 1 year for all couples in

201   the validation cohort using the formula in the Appendix of the paper by van Eekelen *et al* (van

202   Eekelen *et al.*, 2017a). To test the model's ability to not only predict after the completion of

203   the fertility workup, but also when a couple returns after an unsuccessful period of expectant

204   management, we calculated the prognosis at four time points: directly after completion of the

205   workup, after 0.5 year, 1 year and after 1.5 years of expectant management. We evaluated

206   model performance in terms of calibration, i.e. the degree of agreement between observed

207   and predicted natural conception rates, and discrimination, i.e. the ability of the dynamic

208   prediction model to distinguish between couples who do conceive and couples who do not

209   conceive.

210

211   To assess calibration, we first explored whether the overall prediction of the model was

212   correct by comparing the average predicted probability over a time period with the observed

213   conception rate over that same time period. This is referred to as calibration-in-the-large and

214   assesses whether the model systematically under- or overestimates the observed conception

215   rate (Steyerberg, 2009).

216   Second, we assessed whether the effects of patient characteristics were estimated

217   correctly in three ways: by visuals using calibration plots for risk groups, by calibration within

218   groups with similar patient characteristics and by calculating a calibration slope. For the

219   calibration plots we ordered the predicted probabilities of couples and divided them in risk

220   groups with similar predictions (n=135 per risk group). We compared the mean predicted

221   chances within these groups with the corresponding observed fraction of ongoing pregnancy

222   as estimated by the Kaplan-Meier method. We visualized the observed fractions and

223   predicted probabilities per risk group in plots and tabulated the absolute differences. In the

224   plots, the 45 degree line indicates what would be a perfect agreement between the observed

225   fraction and average predicted probability within a risk group.

226   We repeated the calibration procedure but instead of grouping based on predicted risks, we

227   grouped couples based on having similar patient characteristics. We again compared the

228   mean predicted chances within these groups with the corresponding observed fraction of

229   ongoing pregnancy as estimated by the Kaplan-Meier method and tabulated the results.

230   To calculate the calibration slope, we used the prognostic index (i.e. the sum of the

231   multiplication between all patient characteristics and the coefficients from the model) as an

232   explanatory variable in a Cox model for each of the four evaluated time periods (van

233   Houwelingen, 2000). Ideally, the calibration slope is unity i.e. 1, indicating that the strength of

234   the patient characteristics in the evaluated model perfectly matches the validation data.

235   Third, we used a recalibration procedure as an alternative way to assess the

236   systematic under- or overestimation (calibration-in-the-large) and the strength of the patient

237   characteristics (calibration slope) in the model. We did this by using the same coefficients for

238   the patient characteristics as reported by van Eekelen et al. to calculate a prognostic index,

239   but re-estimated the other parameters of the beta-geometric model in the validation dataset

240   (Bongaarts, 1975; Weinberg and Gladen, 1986). The recalibration model re-estimates three

241   parameters, which we compared to those in the van Eekelen model and tested for the

242   difference between the two using independent samples z-tests. Systematic under- or

243   overestimation was assessed by comparing the intercept and the variance parameters. The

244   intercept parameter indicates the estimated pregnancy chances in the first cycle after the

245   fertility workup and the variance parameter indicates how fast the estimated chances

246   decrease over consecutive failed natural cycles. Similarity in strength of the patient

247   characteristics was assessed by again calculating a calibration slope parameter, which would

248   ideally be 1.

249

250   We assessed discrimination by calculating Harrel's c statistic at the four time points, which

251   we compared to those found at internal validation (Harrell *et al.*, 1996).

252   Finally, we explored the range of predicted probabilities at the four time points to see if they

253   facilitate meaningful prognostic stratification of couples (Coppus *et al.*, 2009).

254   All analyses were conducted in R version 3.4.3 and RStudio (R Core Team, 2013). A p value

255   below 0.05 was considered statistically significant.

256

257

258   **Results**

259   Data of 1203 couples were included (Fig. 1). The baseline characteristics of the couples are

260   shown in Table I.

261   In total, 398 (33%) couples conceived naturally, leading to an ongoing pregnancy. The

262   median follow up was 1 year and 3 months after completion of the workup (average follow up

263   2 years and 6 months). The observed rates of natural conception up to 2.5 years are

264   depicted in Fig. 2 (upper panel). For couples who did not yet conceive after 0.5 year, 1 year

265   or 1.5 years after completion of the fertility workup, the observed rates of natural conception

266   over the following year are depicted in Fig. 2 (lower panel). The mean probability of natural

267   conception as predicted by the dynamic model over the course of the first year after the

268   fertility workup was 25% while the observed fraction was 23% (95%CI 20-25). For couples

269   who did not conceive after 0.5 years, after 1 year and after 1.5 years of expectant

270   management, the mean estimated probability of conceiving over the course of the following

271   year was estimated at 18%, 14% and 12%. The observed rates were 15% (13-18%), 14%

272   (11-17%) and 12% (9-15%) for these three time periods, respectively (Fig. 2, lower panel).

273   Except for the second period during which the model slightly overestimated the pregnancy

274   chances by 3 percentage points, the mean predicted probabilities fell within their respective

275   confidence limits of the observed rates, indicating good agreement between the average

276   prediction rendered by the dynamic model and the corresponding observed rate of natural

277   conception.

278

279   The calibration plots for the four time periods are presented in Fig. 3. The dynamic prediction

280   model was well calibrated based on the upward trends observed in the four plots, indicating

281   that higher predicted probabilities correspond to higher observed rates, and the CIs from the

282   observed rates which all but one cover the ideal 45 degree line. The second calibration plot

283   starting at 0.5 years after the fertility workup showed a slight overestimation since all points

284   are below the 45 degree line. The absolute differences between observed fractions and

285   predicted probabilities of natural conception within risk groups are shown in Table II. This

286   was on average 2.8 percentage points and 9.6 at the highest.

287   The results for the calibration grouping couples by similar characteristics are shown in

288   Supplementary Data I. Results were similar to those in the calibration using risk groups, with

289   a slight overestimation in the time periods right after completion of the fertility workup and

290   after 0.5 years of expectant management.

291   The calibration slopes using Cox models were 0.86, 1.01, 1.01 and 0.62 for the four time

292   periods, respectively. None of the corresponding p-values were below 0.05, indicating no

293   statistical evidence for under- or overfitting.

294   In the recalibration model, the intercept and variance parameters were similar to those

295   reported by van Eekelen *et al*. (p=0.69 and p=0.29 for the difference, respectively), indicating

296   similar underlying chances of pregnancy in the first cycle after the workup and a similar

297   decrease in chances as time progresses. The slope was 0.90 (p=0.37), indicating a similar

298   strength of patient characteristics in the validation cohort and no significant difference from 1.

299

300   The discriminative ability of the model in the validation cohort was moderate and similar to

301   that in the Dutch development cohort, ranging over time from a c statistic of 0.61 (95%CI

302   0.57-0.64) in the first year, 0.62 (95% CI 0.58-0.67) from 0.5 years, 0.63 (95% CI 0.57-0.69)

303   from 1 year, to 0.60 (95% CI 0.52-0.67) for 1.5 years after completion of the fertility workup,

304   all for conceiving in the following year. The c statistics were around 0.61 for all four time

305   periods and seemed stable over time.

306

307   The range of predictions varied between 0 and 55% over the course of the first year after the

308   fertility workup. After 0.5 years, 1 year and 1.5 years of expectant management the ranges

309   narrowed to 0 to 43%, 0 to 34% and 0 to 29% respectively, all over the course of the

310   following year, facilitating a distinction between couples with a good or poor prognosis.

311

312   *Sensitivity analyses*

313   Results from the two sensitivity analyses are reported online as supplementary data. The

314   analysis where we assumed 1.5 months between registration and completion of the fertility

315   workup showed a very good performance of the dynamic prediction model (Supplementary

316   Table SI, Supplementary Figs. S1 and S2). The analysis assuming 4.5 months between

317   registration and completion of the fertility workup showed similar results to the primary

318   analysis but with slightly more overestimation of chances by the model (Supplementary

319   Table SII, Supplementary Figs. S3 and S4).

320

321

322   **Discussion**

323   We conducted an external, geographical validation of the van Eekelen model that can be

324   used for repeated predictions of natural conception when couples return to the clinic after

325   unsuccessful expectant management. The model performed well in a Scottish cohort of

326   couples with unexplained subfertility that visited a fertility clinic and the model is expected to

327   be generalizable to other fertility centres and countries where the procedure of managing

328   unexplained subfertile couples is comparable to the Netherlands and the UK. In addition, the

329   predicted probabilities varied sufficiently to aid in distinguishing between couples with a good

330   and poor prognosis in terms of natural conception.

331

332   The data from the AFC was of high quality, registering every unexplained subfertile couple in

333   the Grampian region. All natural conceptions leading to ongoing pregnancy, including after

334   miscarriages and other early pregnancy outcomes, were found using data linkage with

335   maternity records. Indications for the fertility workup and definitions of censoring and

336   prognostic characteristics in the Scottish cohort were very similar to the Dutch cohort, aiding

337   comparability (van Eekelen *et al.*, 2017a).

338   The model was well calibrated, which we consider of higher importance than

339   discrimination since the c statistic can be expected to be moderate due to the limited range

340   of predicted chances in fertility (Mol *et al.*, 2005; Cook, 2007). This restricts the maximum

341   possible c statistic, even if a model were to produce perfect predictions. Recalibration, in

342   which one or more parameters of the prediction model are updated to accommodate better

343   predictions in a different country or clinical setting, was not necessary since the recalibration

344   model showed similar values for all parameters as observed in the development cohort.

345

346   The main limitation to our study was missing data in terms of dates of completion of the

347   fertility workup and menstrual cycle lengths. Menstrual cycle length was not considered very

348   influential since the estimations of the number of cycles per individual are reasonable

349   approximations due to the narrow range of possible cycle lengths in our selection of

350   unexplained subfertile couples, but we did have to make strong assumptions about the date

351   of completion of the fertility workup. We assumed 3 months between registration and

352   completion of the fertility workup, which resulted in ongoing pregnancies before 3 months

353   after registration being excluded. The 'starting' moment of follow up thus differed from the

354   Dutch development cohort since in the latter, the date of last tubal test was used as the end

355   of the workup. Some Dutch clinics did not conduct a visual test of tubal patency, i.e.

356   laparoscopy or hysterosalpingography after a negative result for the chlamydia antibody test.

357   In those Dutch clinics, the workup was thus considered as complete earlier after registration

358   compared to the AFC where visual tests of tubal patency are part of the standard protocol.

359   This may have led to the observed slight overestimation in the first year after the fertility

360   workup and after 0.5 years of expectant management but, despite these differences, the

361   dynamic model was still able to estimate a prognosis that was reasonably accurate on cohort

362   and risk group level. The results from the sensitivity analysis assuming 1.5 months between

363   registration and completion of the fertility workup were very good because the resulting

364   population more closely resembled that of the Dutch development cohort in which the same

365   average duration was observed between registration and the workup completion.

366   Accordingly, in the analysis assuming 4.5 months between registration and completion of the

367   fertility workup, the performance of the dynamic model was poorer because the populations

368   differed more due to additional selection that occurred.

369

370   The dynamic model is able to reassess the chance of natural conception after any given

371   period of expectant management from the completion of the fertility workup onwards. For

372   example, a couple with 1 year secondary subfertility is referred by a general practitioner to

373   the fertility clinic of which the woman is 33 years old at the completion of the fertility workup

374   and the man has 40% progressive motile sperm. Applying our model gives a predicted 38%

375   chance of natural conception over the first year after the workup and they might be advised

376   expectant management. When the couple returns to the clinic after 10 unsuccessful

377   months/cycles, reapplying the model yields 25% chance over the following year, which is a

378   realistic decrease given they have tried for an additional 10 months. This could be a reason

379   to consider starting treatment.

380

381   Both the Hunault model and the dynamic model performed well in external validations,

382   indicating that the added value of the dynamic model lies in the ability to update predictions

383   at later time points (van Eekelen *et al.*, 2017a). This provides clinicians and patients with

384   information regarding their prognosis of natural conception not only right after completion of

385   the fertility workup, but also when the couple returns after an additional, unsuccessful period

386   of expectant management, thus aiding in making clinical decisions at multiple time points

387 throughout a couple's trajectory. The ability to update predictions also aids in studies which

388 include the prognosis of natural conception as an in- or exclusion criterion, since the

389 prognosis of couples who return after unsuccessful expectant management can be updated

390 accurately, leading to the desired homogeneity of the study sample (van den Boogaard *et al.*,

391 2014). The dynamic model is flexible and can be used to predict over any desired number of

392 menstrual cycles, for instance when the couple is interested in time periods shorter or longer

393 than 1 year. In short, the dynamic model has a wider clinical applicability than the Hunault

394 model and should be the model of choice.

395

396 **Conclusion**

397 The van Eekelen model is a valid and robust tool that is ready to use in clinical practice to

398 counsel couples with unexplained subfertility on their individualised chances of natural

399 conception at various points in time, notably when couples return to the clinic after a period

400 of unsuccessful expectant management.

401

402   **Supplementary data**

403   Supplementary data are available at *Human Reproduction* online.

404

405   **Acknowledgements**

414

415   **Authors' roles**

416   NvG, MDJ, BS, FvdV, MvW and MJE conceived the study. MDJ performed the data linkage,

417   storage in the Safe Haven and cleaned the data. RvE, NvG and MJE designed the statistical

418   analysis plan. RvE, MDJ and NvG analysed the data. RvE drafted the manuscript. All authors

419   contributed critical revision to the paper and approved the final manuscript.

420

421   **Funding**

427

428   **Conflicts of interest**

429   None.

430   **References**

431

432   Aboulghar M, Baird D, Collin J, Evers J, Fauser B, Lambalk C and Other A. Intrauterine

433   insemination. *Hum Reprod Update* 2009;**15**:265-277.

434

435   ASRM. Revised American Society for Reproductive Medicine classification of endometriosis:

436   1996. *Fertil Steril* 1997;**67**:817-821.

437

438   Bongaarts J. A method for the estimation of fecundability. *Demography* 1975;**12**:645-660.

439

440   Brandes M, Hamilton CJ, de Bruin JP, Nelen WL and Kremer JA. The relative contribution of

441   IVF to the total ongoing pregnancy rate in a subfertile cohort. *Hum Reprod* 2010;**25**:118-126.

442

443   Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction.

444   *Circulation* 2007;**115**:928-935.

445

446   Coppus SF, van der Veen F, Opmeer BC, Mol BW and Bossuyt PM. Evaluating prediction

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

448

449   Gnoth C, Godehardt D, Godehardt E, Frank-Herrmann P and Freundl G. Time to pregnancy:

450   results of the German prospective study and impact on the management of infertility. *Hum*

451   *Reprod* 2003;**18**:1959-1966.

452

453   Harrell FE, Jr., Lee KL and Mark DB. Multivariable prognostic models: issues in developing

454   models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat*

455   *Med* 1996;**15**:361-387.

456

457   Hunault CC, Habbema JD, Eijkemans MJ, Collins JA, Evers JL and te Velde ER. Two new

458   prediction rules for spontaneous pregnancy leading to live birth among subfertile couples,

459   based on the synthesis of three previous models. *Hum Reprod* 2004;**19**:2019-2026.

460

461   Leushuis E, van der Steeg JW, Steures P, Bossuyt PM, Eijkemans MJ, van der Veen F, Mol

462   BW and Hompes PG. Prediction models in reproductive medicine: a critical appraisal. *Hum*

463   *Reprod Update* 2009;**15**:537-552.

464

465   Mol BW, Coppus SF, Van der Veen F and Bossuyt P. ROC-curves are misleading;

466   calibration is not! *Fertil Steril* 2005;**84**:253-254.

467

468   NVOG, Dutch Society for Obstetrics and Gynaecology. Guideline on: subfertility (2010).

469   Accessed on: 5th of February, 2017. Available from: http://bit.ly/1UhuYMV.

470

471   Pandey S, McLernon DJ, Scotland G, Mollison J, Wordsworth S and Bhattacharya S. Cost of

472   fertility treatment and live birth outcome in women of different ages and BMI. *Hum Reprod*

473   2014;**29**:2199-2211.

474

475   Pandian Z, Gibreel A and Bhattacharya S. In vitro fertilisation for unexplained subfertility.

476   *Cochrane Database Syst Rev* 2015;Cd003357.

477

478   R Core Team (2013). R: A language and environment for statistical computing. R Foundation

479   for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

480

481   Steyerberg E. Clinical Prediction Models: A Practical Approach to Development, Validation

482   and Updating. Springer 2009.

483

484   Tjon-Kon-Fat RI, Bensdorp AJ, Scholten I, Repping S, van Wely M, Mol BW and van der

485   Veen F. IUI and IVF for unexplained subfertility: where did we go wrong? *Hum Reprod*

486   2016;**31**:2665-2667.

487

488   van den Boogaard NM, Bensdorp AJ, Oude Rengerink K, Barnhart K, Bhattacharya S,

489   Custers IM, Coutifaris C, Goverde AJ, Guzick DS, Hughes EC*, et al.* Prognostic profiles and

490   the effectiveness of assisted conception: secondary analyses of individual patient data. *Hum*

491   *Reprod Update* 2014;**20**:141-151.

492

493   a. van Eekelen R, Scholten I, Tjon-Kon-Fat RI, van der Steeg JW, Steures P, Hompes P, van

494   Wely M, van der Veen F, Mol BW, Eijkemans MJ*, et al.* Natural conception: repeated

495   predictions over time. *Hum Reprod* 2017a;**32**:346-353.

496

497   b. van Eekelen R, van Geloven N, van Wely M, McLernon DJ, Eijkemans MJ, Repping S,

498   Steyerberg EW, Mol BW, Bhattacharya S and van der Veen F. Constructing the crystal ball:

499   how to get reliable prognostic information for the management of subfertile couples. *Hum*

500   *Reprod* 2017b;**32**:2153-2158.

501

502   van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival

503   models. *Stat Med* 2000;**19**:3401-3415.

504

505   van der Steeg JW, Steures P, Eijkemans MJ, Habbema JD, Hompes PG, Broekmans FJ,

506   van Dessel HJ, Bossuyt PM, van der Veen F and Mol BW. Pregnancy is predictable: a large-

507   scale prospective external validation of the prediction of spontaneous pregnancy in subfertile

508   couples. *Hum Reprod* 2007;**22**:536-542.

509

510   Veltman-Verhulst SM, Hughes E, Ayeleke RO and Cohlen BJ. Intra-uterine insemination for

511   unexplained subfertility. *Cochrane Database Syst Rev* 2016;**2**:Cd001838.

512

513　Wang X, Chen C, Wang L, Chen D, Guang W and French J. Conception, early pregnancy

514　loss, and time to clinical pregnancy: a population-based prospective study. *Fertil Steril*

515　2003;**79**:577-584.

516

517　Weinberg CR and Gladen BC. The beta-geometric distribution applied to comparative

518　fecundability studies. *Biometrics* 1986;**42**:547-560.

519

520　WHO, World Health Organization. Laboratory Manual for the Examination of Human Semen

521　and Sperm-Cervical Mucus Interaction. 4th edn 1999. Cambridge University Press,

522　Cambridge.

523

524　WHO, World Health Organisation. Laboratory Manual for the Examination and Processing of

525　Human Semen, 5th ed 2010. Geneva: World Health Organization.

526

527　**FIGURE LEGENDS**

528

529　**Figure 1** Flow chart of couples with unexplained subfertility who were considered for

530　inclusion in the external validation.

531

532　**Figure 2** Cumulative chances of natural conception leading to ongoing pregnancy.

533　Cumulative chances after completion of fertility workup (upper panel) and updated chances

534　of natural conception over the course of 1 year at completion of the fertility workup or 0.5

535　years, 1 year and 1.5 years thereafter (lower panel) in the validation cohort.

536　Percentages are Kaplan-Meier estimates of the observed fraction of natural conception

537　leading to ongoing pregnancy.

538

539   **Figure 3** Calibration of the predictions of the dynamic prediction model: predicted versus

540   observed 1 year natural conception rates at four fixed time points.

541

542   **Supplementary Figure S1** Cumulative chances of natural conception leading to ongoing

543   pregnancy after completion of fertility workup (upper panel) and updated chances of natural

544   conception over the course of 1 year at completion of the fertility workup or 0.5 years, 1 year

545   and 1.5 years thereafter (lower panel) in the validation cohort. Percentages are Kaplan-Meier

546   estimates of the observed fraction of natural conception leading to ongoing pregnancy. Data

547   analysis assumed 1.5 months between registration at the Aberdeen Fertility Clinic and

548   completion of the fertility workup (n=1261).

549

550   **Supplementary Figure S2** Calibration of the predictions of the dynamic prediction model:

551   predicted versus observed 1 year natural conception rates at four fixed time points. Data

552   analysis assumed 1.5 months between registration at the Aberdeen Fertility Clinic and

553   completion of the fertility workup (n=1261).
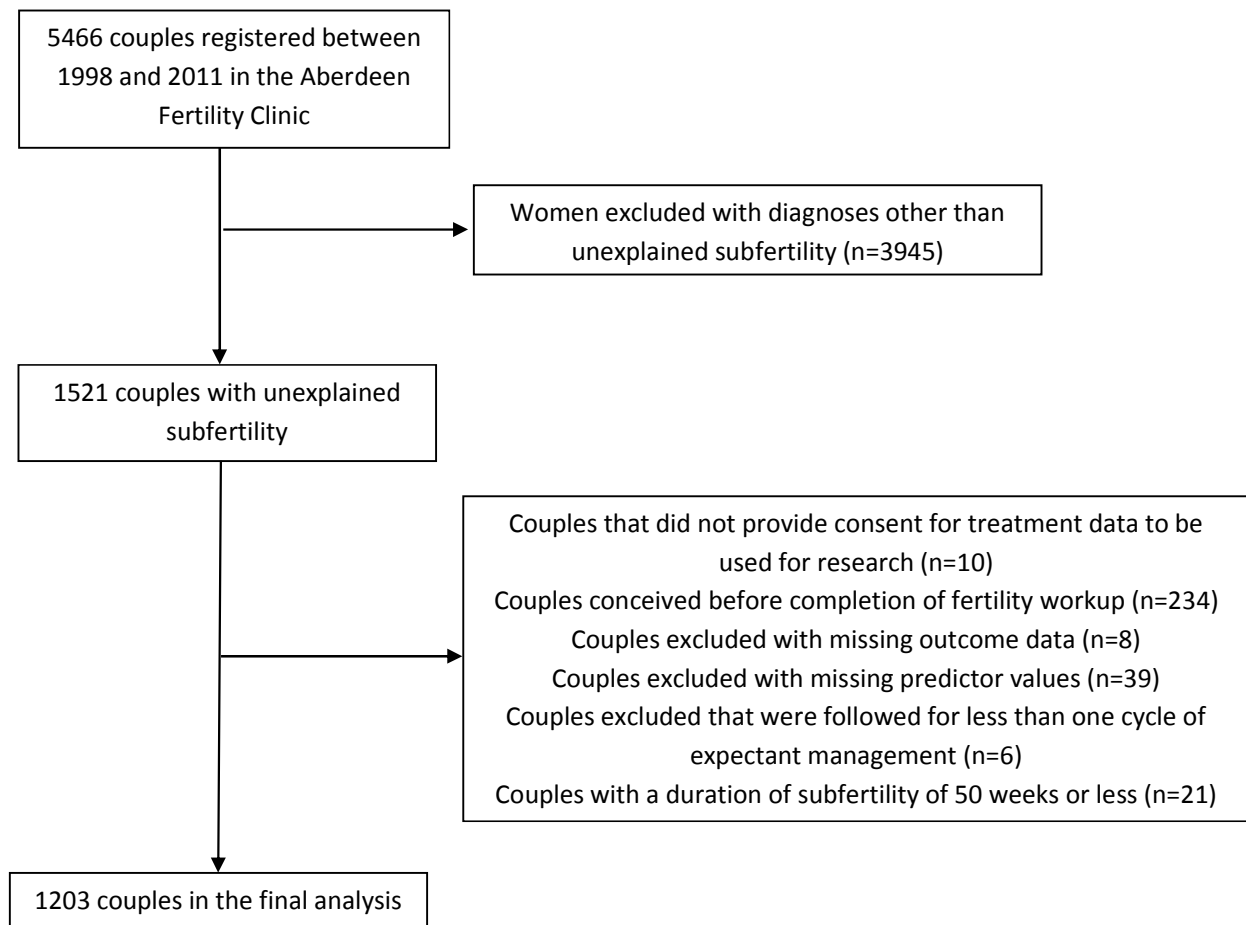
554

555   **Supplementary Figure S3** Cumulative chances of natural conception leading to ongoing

556   pregnancy after completion of fertility workup (upper panel) and updated chances of natural

557   conception over the course of 1 year at completion of the fertility workup or 0.5 years, 1 year

558   and 1.5 years thereafter (lower panel) in the validation cohort. Percentages are Kaplan-Meier

559   estimates of the observed fraction of natural conception leading to ongoing pregnancy. Data

560   analysis assumed 4.5 months between registration at the Aberdeen Fertility Clinic and

561   completion of the fertility workup (n=1123).

562

563

564   **Supplementary Figure S4** Calibration of the predictions of the dynamic prediction model:

565   predicted versus observed 1 year natural conception rates at four fixed time points. Data

566   analysis assumed 4.5 months between registration at the Aberdeen Fertility Clinic and
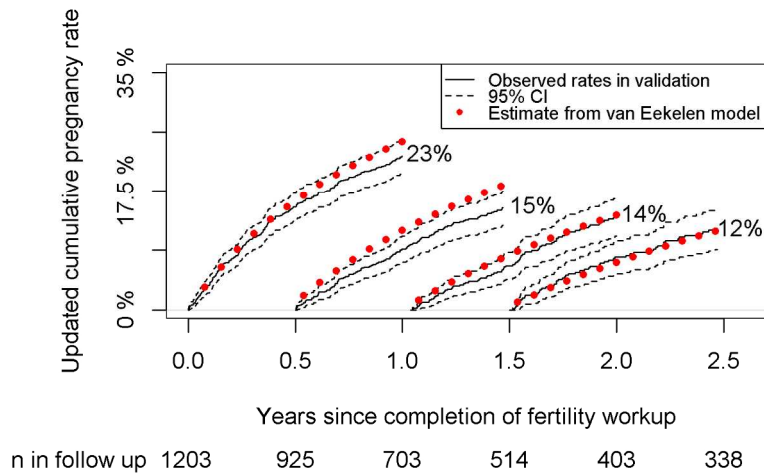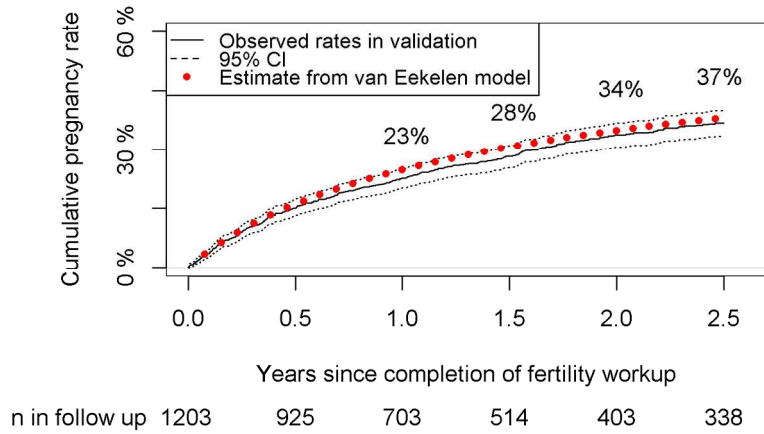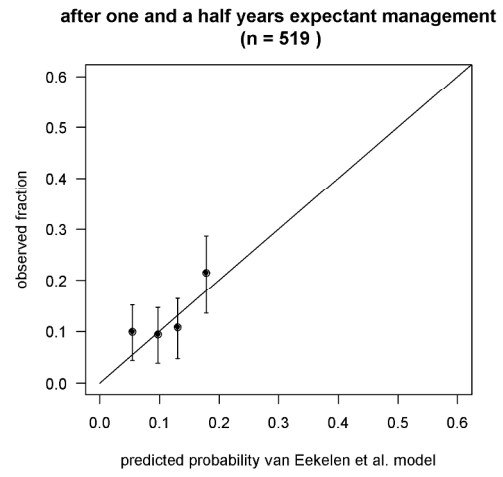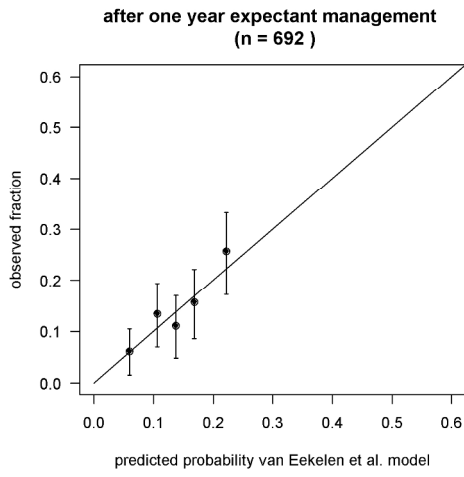
567   completion of the fertility workup (n=1123).

Figure 1.

160x220mm (300 x 300 DPI)

at completion of the fertility workup (n = 1203)

after half a year expectant management (n = 933)

after one year expectant management (n = 692)

after one and a half years expectant management (n = 519)
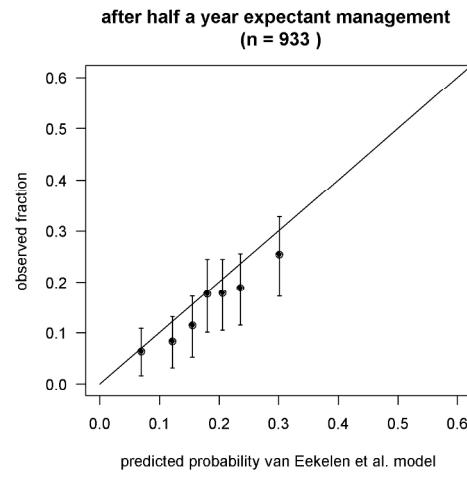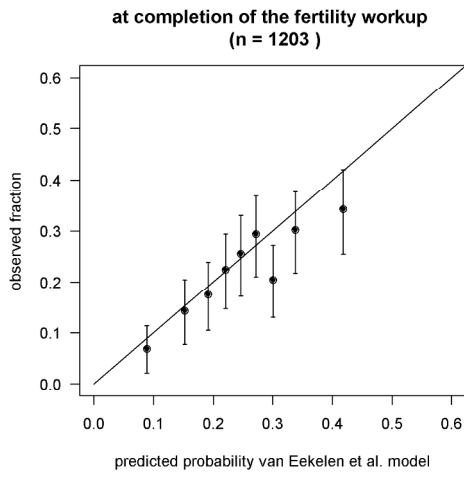
228x228mm (300 x 300 DPI)

Table I Baseline characteristics at completion of the fertility workup.

n = 1203	Mean or n	5th – 95th percentile or %
Female age, in years	33.3	25 - 41
Duration of subfertility, in years	2.7	1.3 - 5.6
Primary female subfertility	697	58%
Percentage of progressive motile sperm	51	24 - 76
Referral by secondary care	84	7%

Table II Calibration of the dynamic prediction model by risk groups.

	Mean difference	Max difference	Number of risk groups
After completion of workup	3.2	9.6	9
After 0.5 year EM	3.0	4.7	7
After 1 year EM	2.1	3.5	5
After 1.5 years EM	2.7	4.5	4
Total	2.8	9.6	25

Data are the mean and maximum of the absolute differences (in percentage points) between predicted and observed 1 year natural conception rates per risk group of n=135, stratified by the elapsed period of expectant management (EM).

Supplementary Data

Calibration per strata of patient characteristics.

Analyses were conducted for all four time periods in the primary scenario with 3 months between registration and completion of fertility workup.

Time period 1: after completion of the fertility workup

Female age:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
<= 28 years [25.4, n=155]	33	30 (22-37)
28-32 years [29.3, n=336]	29	26 (21-30)
32-35 years [33.2, n=261]	24	23 (17-28)
> 35 years [38.3, n=451]	19	18 (14-21)

Duration of subfertility:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
1 - 1.5 years [1.3, n=227]	30	26 (20-32)
1.5-2 years [1.8, n=262]	29	26 (21-32)
2-3 years [2.4, n=393]	25	24 (20-29)
> 3 years [4.7, n=321]	16	14 (10-19)

Percentage of progressive motile sperm:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
<= 35% [26, n=220]	21	18 (13-23)
35-50% [43, n=376]	24	23 (19-28)
50-65% [58, n=366]	26	23 (18-27)
> 65% [73, n=241]	28	26 (19-31)

Primary or secondary subfertility:

Category [n]	Predicted probability in %	Observed fraction in % (95%CI)
Primary [n=697]	23	20 (17-23)
Secondary [n=506]	28	26 (22-30)

Referral by general practitioner (GP) or specialist/gynaecologist:

Category [n]	Predicted probability in %	Observed fraction in % (95%CI)
GP [n=1119]	26	23 (20-25)
Specialist/gynaecologist	13	20 (10-29)

[n=84]		

Time period 2: after 0.5 years of expectant management

Female age:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
<= 28 years [25.4, n=117]	24	22 (13-29)
28-32 years [29.9, n=262]	22	18 (12-22)
32-35 years [33.2, n=205]	17	16 (11-22)
> 35 years [38.3, n=349]	14	11 (7-14)

Duration of subfertility:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
1 - 1.5 years [1.3, n=170]	23	18 (12-24)
1.5-2 years [1.8, n=204]	22	19 (13-25)
2-3 years [2.4, n=309]	19	17 (13-22)
> 3 years [4.7, n=250]	12	8 (4-11)

Percentage of progressive motile sperm:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
<= 35% [26, n=176]	15	14 (8-20)
35-50% [43, n=282]	18	14 (9-18)
50-65% [58, n=291]	19	16 (11-20)
> 65% [73, n=184]	21	18 (12-24)

Primary or secondary subfertility:

Category [n]	Predicted probability in %	Observed fraction in % (95%CI)
Primary [n=550]	17	14 (11-17)
Secondary [n=383]	20	17 (13-21)

Referral by general practitioner (GP) or specialist/gynaecologist:

Category [n]	Predicted probability in %	Observed fraction in % (95%CI)
GP [n=867]	19	15 (13-18)
Specialist/gynaecologist [n=66]	10	14 (4-23)

Time period 3: after 1 year of expectant management

Female age:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
<= 28 years [25.4, n=92]	19	24 (14-34)
28-32 years [29.9, n=189]	17	19 (12-25)
32-35 years [33.2, n=154]	14	15 (8-21)
> 35 years [38.5, n=257]	10	7 (4-10)

Duration of subfertility:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
1 - 1.5 years [1.3, n=129]	18	16 (9-23
1.5-2 years [1.8, n=150]	17	20 (12-27)
2-3 years [2.4, n=223]	14	15 (9-20)
> 3 years [4.8, n=190]	9	9 (5-14)

Percentage of progressive motile sperm:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
<= 35% [26, n=137]	12	16 (9-22)
35-50% [43, n=208]	14	14 (8-19)
50-65% [58, n=211]	15	14 (9-19)
> 65% [73, n=136]	15	15 (8-21)

Primary or secondary subfertility:

Category [n]	Predicted probability in %	Observed fraction in % (95%CI)
Primary [n=402]	13	15 (11-19)
Secondary [n=290]	15	13 (9-17)

Referral by general practitioner (GP) or specialist/gynaecologist:

Category [n]	Predicted probability in %	Observed fraction in % (95%CI)
GP [n=645]	14	15 (12-18)
Specialist/gynaecologist [n=47]	7	11 (1-20)

Time period 4: after 1.5 years of expectant management

Female age:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
<= 28 years [25.3, n=69]	16	21 (10-30)
28-32 years [29.9, n=143]	14	21 (13-28)
32-35 years [33.2, n=107]	11	11 (4-17)
> 35 years [38.8, n=200]	9	6 (2-9)

Duration of subfertility:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
1 - 1.5 years [1.3, n=93]	15	10 (3-16)
1.5-2 years [1.8, n=104]	14	19 (11-26)
2-3 years [2.4, n=172]	12	12 (6-17)
> 3 years [4.8, n=150]	8	12 (6-17)

Percentage of progressive motile sperm:

Category [mean, n]	Predicted probability in %	Observed fraction in % (95%CI)
<= 35% [25, n=95]	10	14 (6-21)
35-50% [43, n=161]	12	13 (7-19)
50-65% [58, n=161]	12	14 (8-19)
> 65% [73, n=102]	13	11 (4-17)

Primary or secondary subfertility:

Category [n]	Predicted probability in %	Observed fraction in % (95%CI)
Primary [n=286]	11	14 (9-18)
Secondary [n=233]	13	12 (7-16)

Referral by general practitioner (GP) or specialist/gynaecologist:

| Category [n]	Predicted probability in %	Observed fraction in % (95%CI)
GP [n=485]	12	13 (9-16)
Specialist/gynaecologist [n=34]	6	16 (2-29)

1   **Supplementary Table SI** Calibration of the dynamic prediction model by risk groups.*

2

	mean difference	max difference	number of risk groups
After completion of workup	1.9	5.3	9
After 0.5 years EM	1.7	3.4	7
After 1 year EM	1.8	3.0	5
After 1.5 years EM	2.9	4.0	4
Total	2.1	5.3	25

3   *Assuming 1.5 months between registration at the Aberdeen Fertility Clinic and completion of
4   the fertility workup (n=1261).
5   Data are the mean and maximum of the absolute differences (in percentage points) between
6   predicted and observed 1 year natural conception rates per risk group of n=135, stratified by
7   the elapsed period of expectant management (EM).

8   **Supplementary Figure S1** Cumulative chances of natural conception leading to ongoing
9   pregnancy after completion of fertility workup (upper panel) and updated chances of natural
10   conception over the course of 1 year at completion of the fertility workup or 0.5 years, 1 year
11   and 1.5 years thereafter (lower panel) in the validation cohort. Percentages are Kaplan-Meier
12   estimates of the observed fraction of natural conception leading to ongoing pregnancy. Data
13   analysis assumed 1.5 months between registration at the Aberdeen Fertility Clinic and
14   completion of the fertility workup (n=1261).
15
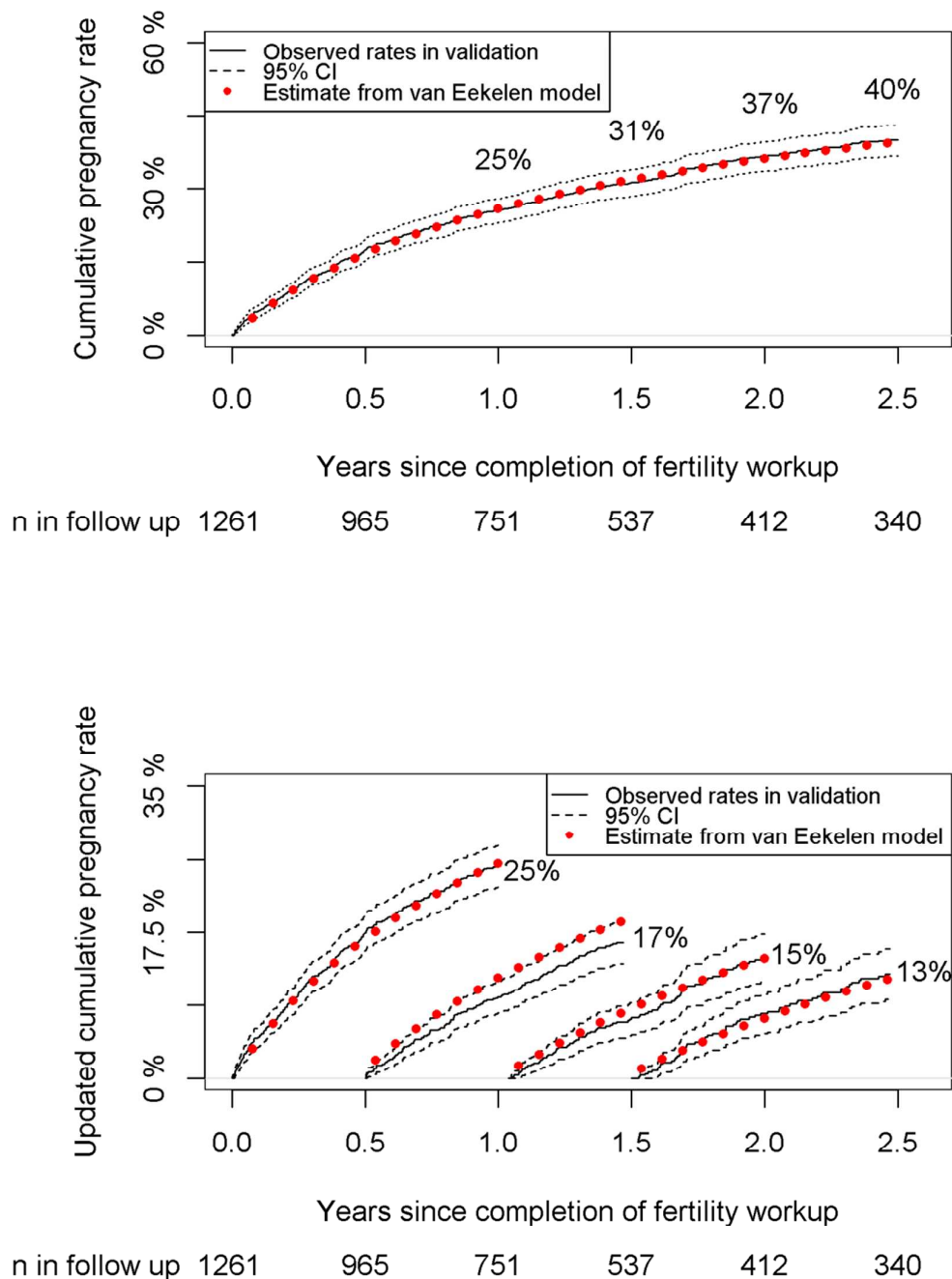
16   **Supplementary Figure S2** Calibration of the predictions of the dynamic prediction model:
17   predicted versus observed 1 year natural conception rates at four fixed time points. Data
18   analysis assumed 1.5 months between registration at the Aberdeen Fertility Clinic and
19   completion of the fertility workup (n=1261).
20
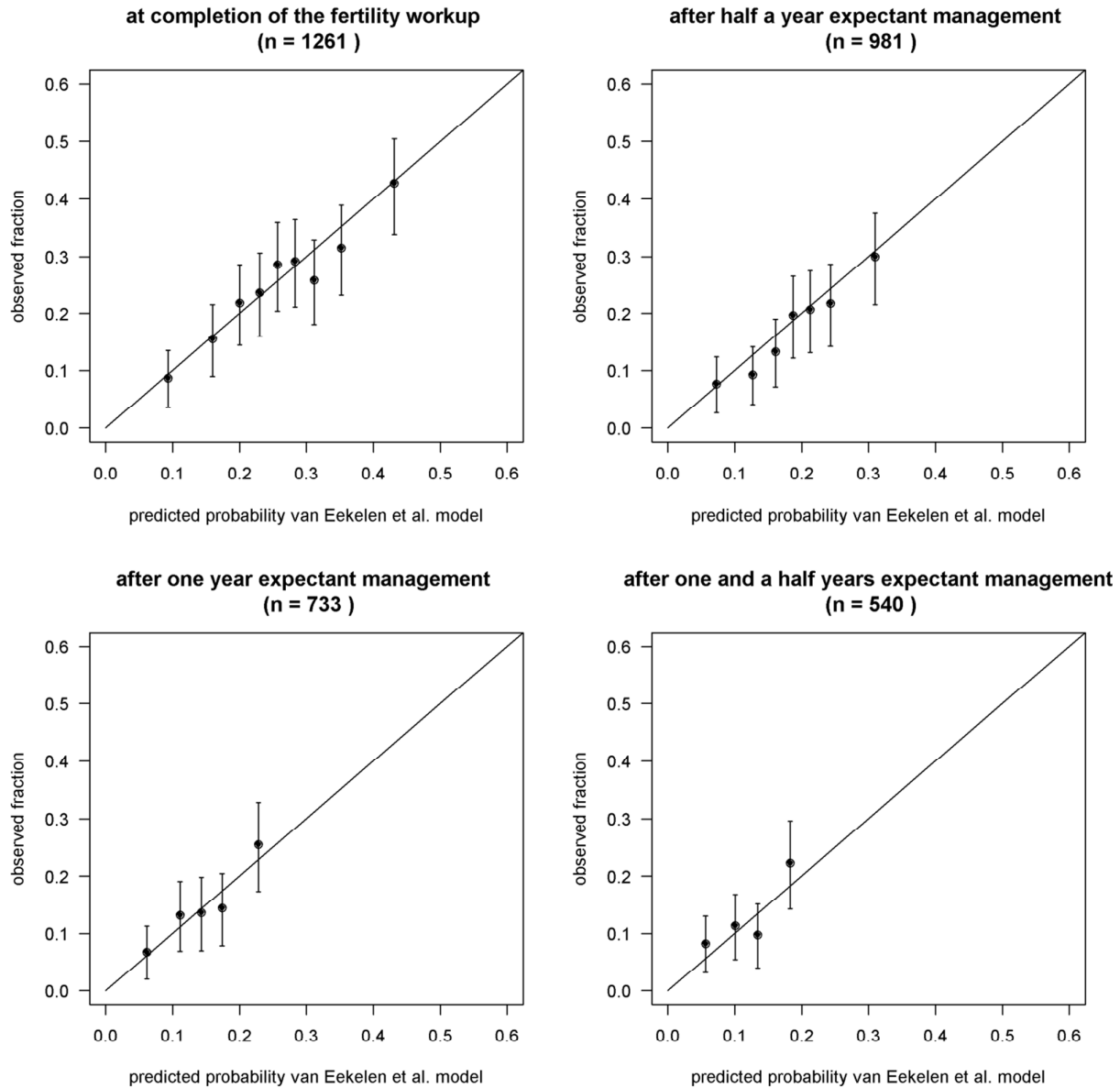
21
22

23   **Supplementary Table SII** Calibration of the dynamic prediction model by risk groups.*

24

	mean difference	max difference	number of risk groups
After completion of workup	3.9	11.1	8
After half a year EM	1.9	5.8	7
After one year EM	2.6	4.3	5
After one and a half years EM	2.3	3.3	4
Total	2.7	11.1	24

25   *Data analysis assumed 4.5 months between registration at the Aberdeen Fertility Clinic and
26   completion of the fertility workup (n=1123).
27   Data are the mean and maximum of the absolute differences (in percentage points) between
28   predicted and observed 1 year natural conception rates per risk group of n=135, stratified by
29   the elapsed period of expectant management (EM).

30　**Supplementary Figure S3** Cumulative chances of natural conception leading to ongoing
31　pregnancy after completion of fertility workup (upper panel) and updated chances of natural
32　conception over the course of 1 year at completion of the fertility workup or 0.5 years, 1 year
33　and 1.5 years thereafter (lower panel) in the validation cohort. Percentages are Kaplan-Meier
34　estimates of the observed fraction of natural conception leading to ongoing pregnancy. Data
35　analysis assumed 4.5 months between registration at the Aberdeen Fertility Clinic and
36　completion of the fertility workup (n=1123).
37

38
39
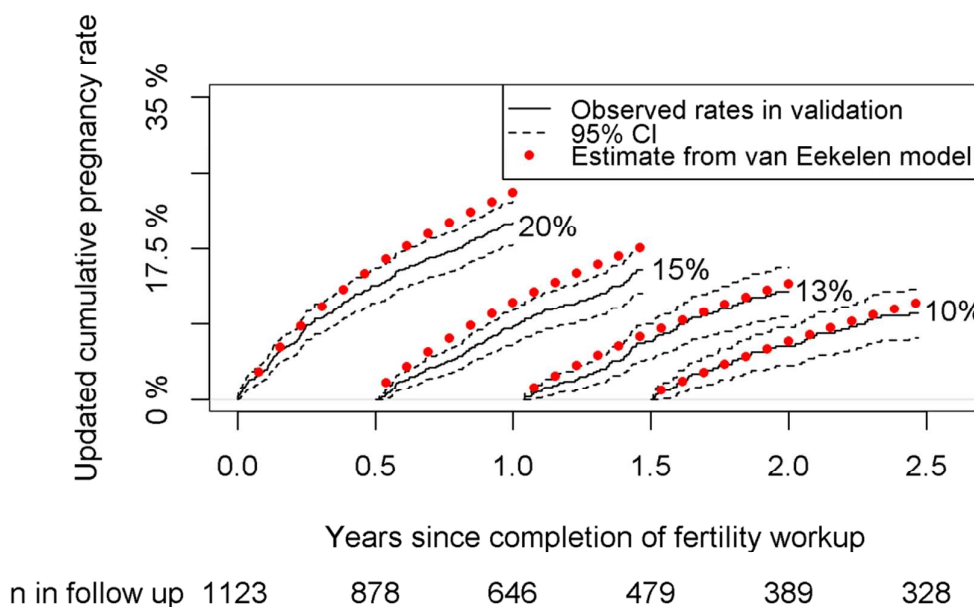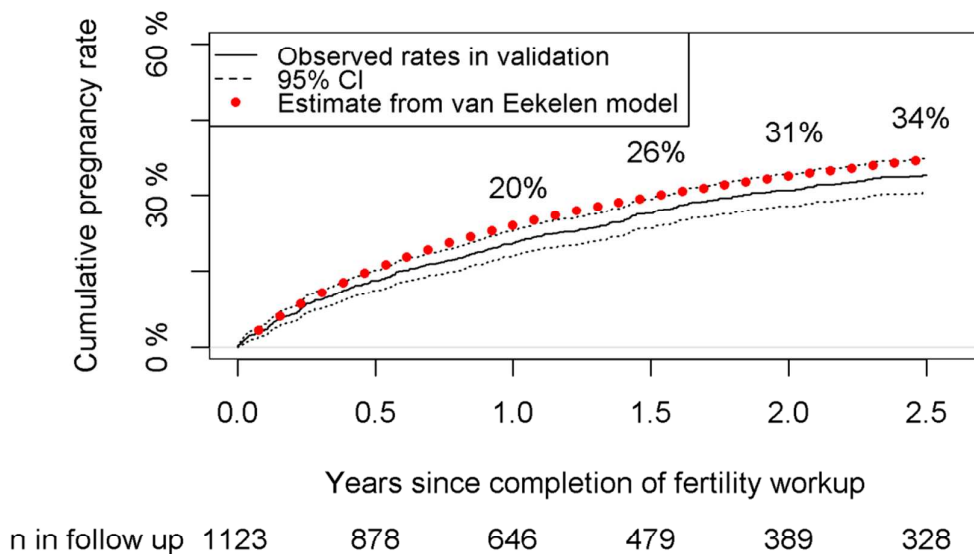
40   **Supplementary Figure S4** Calibration of the predictions of the dynamic prediction model:
41   predicted versus observed 1 year natural conception rates at four fixed time points. Data
42   analysis assumed 4.5 months between registration at the Aberdeen Fertility Clinic and
43   completion of the fertility workup (n=1123).
44