External validation of a frequently used prediction model for ongoing pregnancy in couples with unexplained recurrent pregnancy loss

Youssef, A.; Hoorn, M.L.P. van der; Dongen, M.; Visser, J.; Bloemenkamp, K.; Lith, J. van; ... ; Lashley, E.E.L.O.

Human Reproduction, Vol.37, No.3, pp. 393–399, 2022
Advance Access Publication on December 7, 2021 https://doi.org/10.1093/humrep/deab264

human reproduction

ORIGINAL ARTICLE *Early pregnancy*

External validation of a frequently used prediction model for ongoing pregnancy in couples with unexplained recurrent pregnancy loss

A. Youssef [1,*], M.L.P. van der Hoorn[1], M. Dongen[1], J. Visser[2], K. Bloemenkamp[3], J. van Lith[1], N. van Geloven[4], and E.E.L.O. Lashley[1]

[1]Department of Obstetrics and Gynaecology, Leiden University Medical Centre, Leiden, The Netherlands [2]Department of Obstetrics and Gynaecology, Amphia Hospital, Breda, The Netherlands [3]Department of Obstetrics and Gynaecology, University Medical Centre Utrecht, Utrecht, The Netherlands [4]Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, The Netherlands

*Correspondence address. Department of Obstetrics and Gynaecology, Leiden University Medical Centre, Leiden 2333 ZA, The Netherlands. E-mail: a.youssef@lumc.nl, https://orcid.org/0000-0003-4385-7971

Submitted on April 23, 2021; resubmitted on September 24, 2021; editorial decision on November 02, 2021

STUDY QUESTION: What is the predictive performance of a currently recommended prediction model in an external Dutch cohort of couples with unexplained recurrent pregnancy loss (RPL)?

SUMMARY ANSWER: The model shows poor predictive performance on a new population; it overestimates, predicts too extremely and has a poor discriminative ability.

WHAT IS KNOWN ALREADY: In 50–75% of couples with RPL, no risk factor or cause can be determined and RPL remains unexplained. Clinical management in RPL is primarily focused on providing supportive care, in which counselling on prognosis is a main pillar. A frequently used prediction model for unexplained RPL, developed by Brigham *et al.* in 1999, estimates the chance of a successful pregnancy based on number of previous pregnancy losses and maternal age. This prediction model has never been externally validated.

STUDY DESIGN, SIZE, DURATION: This retrospective cohort study consisted of 739 couples with unexplained RPL who visited the RPL clinic of the Leiden University Medical Centre between 2004 and 2019.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Unexplained RPL was defined as the loss of two or more pregnancies before 24 weeks, without the presence of an identifiable cause for the pregnancy losses, according to the ESHRE guideline. Obstetrical history and maternal age were noted at intake at the RPL clinic. The outcome of the first pregnancy after intake was documented. The performance of Brigham's model was evaluated through calibration and discrimination, in which the predicted pregnancy rates were compared to the observed pregnancy rates.

MAIN RESULTS AND THE ROLE OF CHANCE: The cohort included 739 women with a mean age of 33.1 years (±4.7 years) and with a median of three pregnancy losses at intake (range 2–10). The mean predicted pregnancy success rate was 9.8 percentage points higher in the Brigham model than the observed pregnancy success rate in the dataset (73.9% vs 64.0% (95% CI for the 9.8% difference 6.3–13.3%)). Calibration showed overestimation of the model and too extreme predictions, with a negative calibration intercept of −0.46 (95% CI −0.62 to −0.31) and a calibration slope of 0.42 (95% CI 0.11–0.73). The discriminative ability of the model was very low with a concordance statistic of 0.55 (95% CI 0.51–0.59). Recalibration of the Brigham model hardly improved the *c*-statistic (0.57; 95% CI 0.53–0.62)

LIMITATIONS, REASONS FOR CAUTION: This is a retrospective study in which only the first pregnancy after intake was registered. There was no time frame as inclusion criterium, which is of importance in the counselling of couples with unexplained RPL. Only cases with a known pregnancy outcome were included.

WIDER IMPLICATIONS OF THE FINDINGS: This is the first study externally validating the Brigham prognostic model that estimates the chance of a successful pregnancy in couples with unexplained RPL. The results show that the frequently used model overestimates the

chances of a successful pregnancy, that predictions are too extreme on both the high and low ends and that they are not much more discriminative than random luck. There is a need for revising the prediction model to estimate the chance of a successful pregnancy in couples with unexplained RPL more accurately.

STUDY FUNDING/COMPETING INTEREST(S): No external funding was used and no competing interests were declared.

TRIAL REGISTRATION NUMBER: N/A.

Key word: recurrent pregnancy loss / external validation / prediction model / pregnancy success rate / miscarriage

Introduction

Recurrent pregnancy loss (RPL) is defined as the loss of two or more conceptions (Bender Atik *et al.*, 2018). This condition affects 1–3% of all fertile couples (Jauniaux *et al.*, 2006; Rai and Regan, 2006). RPL is a highly heterogeneous condition with multiple known maternal and paternal risk factors (Nybo Andersen *et al.*, 2004; Venners *et al.*, 2004; McQueen *et al.*, 2019). Despite extensive diagnostic work-ups offered to couples with RPL, an underlying risk factor may be identified in only 25–50% of couples (Stephenson, 1996; Jaslow *et al.*, 2010). Limited understanding of mechanisms underlying RPL leads to the lack of options for effective treatment. As no evidence-based therapeutic options are available for couples with RPL, clinical management is primarily focused on providing supportive care. Supportive care and intensive pregnancy surveillance in the first trimester of gestation are assumed to be of influence in the prevention of new pregnancy loss (Liddell *et al.*, 1991). An important aspect of this supportive care is counselling on the prognosis and success rate of subsequent pregnancies in couples with RPL.

Several prediction models for the estimation of the chance of live birth after RPL have been published (Cauchi *et al.*, 1991, 1995; Quenby and Farquharson, 1993; Brigham *et al.*, 1999; Sugiura-Ogasawara *et al.*, 2009; Lund *et al.*, 2012; Bashiri *et al.*, 2020) and various international guidelines recommend the use of different prediction models (Youssef *et al.*, 2019). The ESHRE RPL guideline recommends to use the prediction models of Brigham *et al.* or Lund *et al.* (hereafter called the 'Brigham model' and the 'Lund model') to estimate the chance of live birth in couples with unexplained RPL (Bender Atik *et al.*, 2018). The Brigham model has been implemented in RPL care in the Netherlands and in the UK (NVOG, 2007; RCOG, 2011), while the American Society for Reproductive Medicine (ASRM) adapted the Lund model in their RPL guideline (Practice Committee of the American Society for Reproductive Medicine, 2012). The Lund model was not designed for individual risk assessment, given the descriptive scope of the study. Furthermore, the study does not discriminate between unexplained and explained RPL. Although the Brigham model and the Lund model were both reviewed with high methodological quality and both studies have consistent results, these models did not follow the nowadays recommended TRIPOD guideline in the development and reporting of a prediction model (Collins *et al.*, 2015). This guideline provides a 22-item checklist consisting of items that assures transparent reporting and acts as a tool for reminding authors of all necessary prediction components, such as measuring the predictive performance of the study internally and/or externally. Both models were never internally nor externally validated, which leaves their predictive performance unknown.

As the Lund model was not intended for individual risk assessment, the aim of this study is to externally validate the Brigham model to assess its predictive performance in a Dutch cohort of couples with unexplained RPL.

Materials and methods

Patient population

We included couples with unexplained RPL who visited the clinic of the Leiden University Medical Centre (LUMC) for intake consultation between 2004 and 2019. We defined unexplained RPL as the loss of two or more pregnancies until 24 weeks, without the presence of an identifiable cause for the pregnancy losses, according to the ESHRE guideline (Bender Atik *et al.*, 2018). The following investigations were performed to rule out factors associated with RPL: maternal testing for antiphospholid syndrome (lupus antibodies, anticardiolipin antibodies, anti-β2-glycoprotein antibodies), parental karyotyping for chromosomal abnormalities based on a priori chance (Franssen *et al.*, 2005), endocrinological factors (thyroid function and thyroid peroxidase antibody testing, random glucose level on indication (Barents, 2021)) and assessment of uterine cavity to rule out anatomic abnormalities. Testing for inherited thrombophilia and hyperhomocysteinemia was performed until 2018 as these were regarded as associated factors for RPL. Since the publication of the ESHRE guideline in November 2017, thrombophilia and hyperhomocysteinemia testing were excluded from the RPL investigations and are only performed to rule out an increased chance of thrombotic events, as is now daily practice at our clinic. RPL couples who tested positive for either, but did not have any other associated RPL factors, were regarded as unexplained RPL in this study. After intake at the LUMC RPL clinic, intensive pregnancy surveillance in the first weeks of gestation was offered in a new pregnancy, consisting of weekly ultrasound checks performed by an easily accessible and dedicated RPL team.

Data collection

Data collection was performed according to the Brigham model. We retrieved maternal age and number of preceding miscarriages at time of intake at the RPL clinic. The outcome of the first pregnancy after intake at the clinic was registered. A successful outcome was regarded as ongoing pregnancy (heartbeat on ultrasound) beyond 24 weeks. Only patients with a known pregnancy outcome were considered for inclusion. Couples missing this data were assumed not to differ systematically from couples with complete data.

Statistical analysis

We evaluated the predictions of the Brigham model through calibration and discrimination. Calibration examines the agreement between

the predicted and observed pregnancy success rates, while discrimination refers to the ability of the model to separate women with a successful pregnancy from those without. Therefore, we calculated the percentages of a successful pregnancy according to the formula described by the Brigham model, as shown below (Brigham *et al.*, 1999).

$$\log\left(\frac{P}{1-P}\right) = 2.00 - 0.0828\ (\text{age} - 32) - 0.2467\ (\text{number of pregnancy losses})$$

Here, P is the predicted probability of a vital pregnancy for those patients who reached pregnancy. We performed a graphical assessment of the calibration, using the *val.prob.ci.2* function, obtained from the library *CalibrationCurves* (https://github.com/BavoDC/CalibrationCurves), of the R statistical program (version 4.0.2). This function validates predicted probabilities against binary events, computing a set of indexes and statistics.

Based on these indexes and statistics, a calibration curve is plotted, including a calibration intercept, which indicates the extent that predictions are systematically too low or too high (also called 'calibration in the large'), and a calibration slope. In a perfectly calibrated model, the intercept is 0 and the slope is 1. An intercept with a negative value suggests overestimation, while an intercept with a positive value suggests underestimation. A slope <1 suggests that the estimated chances are too extreme, while a slope >1 suggests that the estimated risks are too moderate (Van Calster *et al.*, 2019).

The discriminative ability of Brigham's model was measured using the concordance statistic (*c*-statistic). It gives the probability that a randomly selected patient who achieved a successful pregnancy had a higher estimated chance than a patient who did not. A value of 1 means that the model perfectly predicts who will experience a successful pregnancy and who will not. A value of 0.5 means that the model is no better at predicting than random chance.

To see whether the Brigham model would perform better after recalibration to our validation data, we followed the methods described by Vergouwe *et al.* (2017). Three additional logistic regression models were estimated: one updating the intercept of the model (recalibration in the large), one updating the intercept and the strength of the predictors (logistic recalibration) and model revision (estimating all model parameters anew). The performance of these updated models was assessed using the same metrics as for the original Brigham model.

Sample size calculation

For the calculation of the required sample size for this external validation, we used the method described by Riley *et al.* (2020) for the calculation of a sample size in clinical prediction models. We indicated that we were using the same two variables as Brigham: age and number of previous first-trimester pregnancy losses, both as continuous variables. A value of 0.1089 was calculated for the R^2, the expected shrinkage was set to 0.9, as suggested by Riley *et al.* The prevalence of a pregnancy loss was expected to be 35% (Youssef *et al.*, 2020). The R package *pmsampsize* provided alongside the paper of Riley *et al.* was used for the calculation of the sample size. Each step leads to a calculated sample size, and the largest sample size is the required sample size. This resulted in a sample size of 350 couples with unexplained RPL who achieved a new pregnancy after intake at the clinic.

Ethical approval

Approval for this study and data collection was obtained at the Medical Research Ethics Committee of the LUMC (protocols P11.196 and P19.014).

Results

Between 2004 and 2019, 904 couples with unexplained RPL were registered at the LUMC RPL clinic (Supplementary Fig. S1). Of these 904 couples, 107 (11.8%) were lost to follow-up, and 58 couples did not conceive a pregnancy after intake, which resulted in a group of 739 couples with a known outcome of the first pregnancy after intake at the RPL clinic. These 739 couples are included in the analysis (Fig. 1). The mean age of the women was 33.1 years (±4.7 years), with a median of three pregnancy losses at intake (range 2–10 pregnancy losses). More than half of the couples (60.5%) had not previously given birth (live births; range 0–4). The baseline characteristics of these couples are

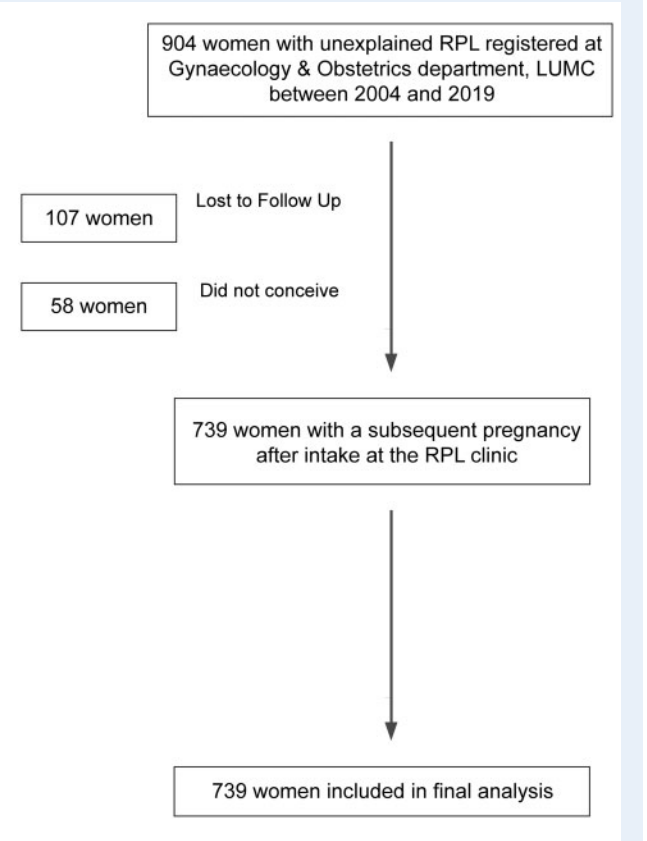

Figure 1. Flow chart of women with unexplained recurrent pregnancy loss who were considered for inclusion in external the validation. Earlier research described a cohort of women visiting the Leiden University Medical Centre recurrent pregnancy loss (RPL) clinic found that 71.5% of women had unexplained RPL (Youssef *et al.*, 2020). The total population with RPL in this cohort would therefore have consisted of about 1264 patients. Supplementary Fig. S1 shows the distribution of RPL associated factors in the previously described cohort.

Table I Baseline characteristics at time of intake (n = 739).

Age (years)	**33.1 (±4.7)***
20–24	35 (4.7%)
25–29	140 (18.9%)
30–34	276 (37.3%)
35–39	234 (31.7%)
≥40	54 (7.3%)
Number of previous pregnancy losses (n)	**3 (2–10)+**
2	103 (13.9%)
3	394 (53.3%)
4	150 (20.3%)
≥5	92 (12.4%)
Previous live birth (n)	**0 (0–4)+**
0	447 (60.5%)
1	236 (31.9%)
≥2	56 (7.6%)
Year of inclusion (n)	
2000–2004	50 (6.8%)
2005–2009	180 (24.4%)
2010–2014	279 (37.8%)
2015–2019	230 (31.1%)

*Mean with standard deviation between parentheses.
+Median with range between parentheses.

Table II Overview of outcome data in numbers (n = 739).

No pregnancy	58 (6.4%)+
Lost to follow–up	107 (11.8%)+
Biochemical pregnancy	74 (10.0%)
Clinical pregnancy loss in first trimester	158 (21.4%)
Clinical pregnancy loss in second trimester	2 (0.3%)
Live birth (pregnancy ≥24 weeks gestation)	474 (64.1%)
Pregnancy loss (not further clarified)	31 (4.2%)

+Percentage calculated based on cohort population before exclusion (n = 904).

shown in Table I. The group of patients who were lost to follow-up was comparable at baseline, with a mean age of 33.6 years (±4.7 years), a median of three pregnancy losses at intake (range 2–8 pregnancy losses) and a median of zero live births (range 0–5). The first pregnancy after intake was successful in 64.1% (95% CI 60.6–67.6%) of the couples, defined as a heartbeat on ultrasound ≥24 weeks pregnancy. Data regarding the first pregnancy after intake are shown in Table II.

We plotted the expected success probabilities of the first pregnancy after intake according to Brigham's formula against the observed rates (Fig. 2). The mean predicted pregnancy success rate using the Brigham model was 9.8 percentage points higher than the observed pregnancy success rate in the dataset (73.9% vs 64.0% (95% CI for the 9.8% difference 6.3–13.3%)).

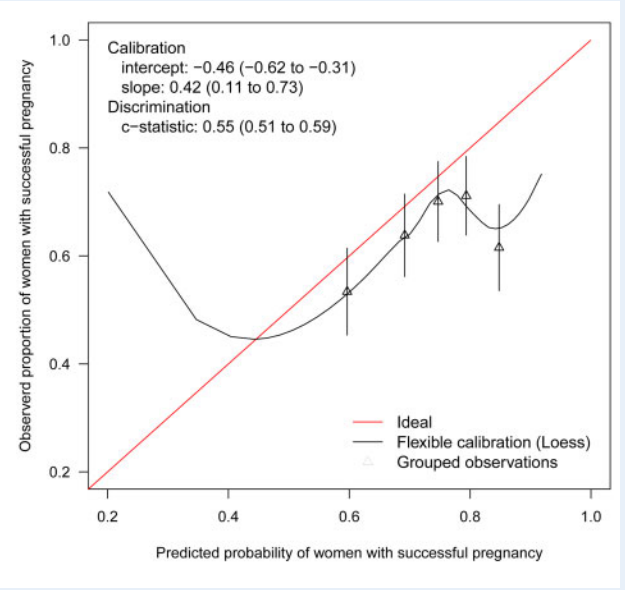

Figure 2. Calibration plot with predicted probabilities according to the Brigham model on the *x*-axis and the observed proportion on the *y*-axis, with a fitted line through the quantiles (n = 739).

Calibration in the large resulted in a statistically significant intercept of −0.46 (95% CI −0.62 to −0.31), affirming the higher predicted success rate. The slope of the calibration curve was statistically significant at 0.42 (95% CI 0.11–0.73). The *c*-statistic, used to describe the discriminative ability of the prediction model, was 0.55 (95% CI 0.51–0.59).

Calibration in the large, logistic recalibration and model revision each led to an improvement in model fit (each Likelihood ratio test comparing against the original model *P*-value < 0.001), thus full model revision was adopted. The revised model was estimated as follows:

$$\log\left(\frac{P}{1-P}\right) = 1.53 - 0.01\ (\text{age} - 32) - 0.28\ (\text{number of pregnancy losses})$$

However, the updated model barely improved its discriminative ability (*c*-statistic 0.57; 95% CI 0.53–0.62).

Discussion

To improve counselling as part of supportive care of RPL couples, accurate predictions on pregnancy success are of utmost importance. This study is the first to externally validate the frequently used Brigham model that predicts the outcome of next pregnancy in couples with unexplained RPL, as developed by Brigham *et al.* (1999). This resulted in a calibration curve with a negative intercept, a slope smaller than 1.0 and a *c*-statistic of 0.55.

A calibration slope of <1 suggests that the estimated risks are too extreme, meaning that the predicted chances are too low for older couples with a higher number of pregnancy losses and that the predicted chances are too high for younger couples with lower number of

pregnancy losses. In other words, the effect of age and number of pregnancy losses is stronger in the Brigham model than in the validation dataset. The value of the *c*-statistic ranges from 0.5 to 1.0, with 0.5 indicating prediction based on pure chance and 1.0 indicating perfect prediction. According to our analysis, there is a poor predictive performance of this model on a new population. The model overestimates, has too extreme predictions and has a poor discriminative ability.

It is already known that the accuracy of prediction models is often lower in a separate cohort (Bleeker *et al.*, 2003). We tried updating the model in our new data, however, the discriminative ability did not improve and the model revision led to re-estimation of all coefficients, which disregards information from the original dataset. Our data suggest that age and number of previous pregnancy losses alone are not able to discriminate between patients with or without a successful next pregnancy.

The ESHRE and RCOG (Royal College of Obstetricians and Gynaecologists) guidelines mention that couples with unexplained RPL have high chances of achieving a live birth in the future, using the Brigham prediction model as substantiation. In our study, however, we observed that the predicted chances of the model are much higher than the actual success rate, reflected by the 9.8 percentage points difference between the mean predicted success rate and the actual live birth rate. The majority (76%) of patients in the dataset of the Brigham model had a history of three or more miscarriages, and the remaining 24% consisted of patients with two miscarriages who requested analysis for the RPL. In our dataset only 14% (103/739) of patients experienced two miscarriages, which could explain the overall lower mean chance of success.

We expected a higher age in our study population, as in general, a trend of delaying motherhood is present (CBS, 2018). This higher age could also explain the observed difference in predicted pregnancy success. However, the mean age in our cohort (33.1 ± 4.8 SD years) does not differ from the mean age in the cohort in the Brigham model (32 years), though for the latter the age range was not presented. Finally, the setting of the two cohorts could be different. Our centre is a tertiary referral centre, but also includes patients referred by primary care. The setting of Brigham's cohort is unknown.

The poor performance of the model in our cohort could also be explained by the model's development. The Brigham model was based on a prospectively collected dataset of 716 patients with RPL. However, only 325 of them were identified as having 'idiopathic recurrent miscarriage' and 23 patients were lost to follow-up. A subsequent pregnancy was achieved by 226/325 (70%) patients, of which two were found to be ectopic, and two patients underwent termination of pregnancy. Thus, the model was based on only 222 patients and this small number could have resulted in overfitting of the model. This is demonstrated in the sample size calculation, which points to a total of 350 patients necessary for a model with two continuous variables. Furthermore, as no internal validation was performed during model development to correct for the degree of overfitting (such as bootstrapping), it is evident that the performance of the model is better on its training dataset than in another or external dataset (Copas, 1983). In short, there was poor development of the study due to underpowering, lack of internal validation and lack of external validation.

Next, the likelihood of finding a low predictive accuracy during validation will increase if a more stringent form of validation is used (Moons *et al.*, 2012a,b). In our study, we included patients from another geographical area and from another time period. This has an influence on differences between the populations. First, the definition of RPL has significantly changed over the past 20 years (Bender Atik *et al.*, 2018). Women with antiphospholipid syndrome, oligomenorrhoea, cervical weakness and abnormal parental chromosome karyotype and patients with a history of second-trimester loss were excluded from the dataset in the Brigham model. According to the current definition, oligomenorrhoea is not considered a factor for RPLs. Furthermore, RPL nowadays includes all pregnancy losses from the time of conception until 24 weeks of gestation. Brigham *et al.* also excluded 'those who had completed successful treatment of an abnormal finding', which is not specified any further in the study.

This study is the first to externally validate the Brigham model, a frequently used prognostic model for successful pregnancy in RPL care. With the large sample size in our study, our evaluation of the model provides precise model performance measures. We followed Brigham's research method to the best of our abilities, to ensure that the external validation was performed on equally developed models. Regarding the outcome, pregnancy success was defined as a pregnancy continuing beyond 24 weeks of gestation, rather than a live birth, which is what patients ultimately want to know. As indicated by Smith *et al.* (2019), there is a need for standardized and patient-central clinical outcomes in studies on pregnancy after RPL.

Importantly, our study only included cases with a known pregnancy outcome in the analysis. In our cohort, the main reason for unknown pregnancy outcomes is that couples leave the clinic around the 10th week of gestation and continue their pregnancy care given by a community midwife. We assumed that missing data was unrelated to the variables involved in the analysis, and therefore did not bias the analysis. This assumption was supported by the fact that patients with missing data were comparable in age, pregnancy losses and live births at baseline. Moreover, missing data and loss of follow-up could also be explained by the inability of couples to achieving a new pregnancy, either voluntary or involuntary, and these couples would not have been included for this study.

Our study shows that the Brigham model does not perform well in a Dutch population. The poor discriminative ability of this model implies that it should not be used routinely in the counselling and prognosis on subsequent pregnancies in patients with RPL. Instead, the model should be revised to estimate the chance of a successful pregnancy in couples with unexplained RPL more accurately.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data underlying this article will be shared on request to the corresponding author.

Authors' roles

A.Y., E.E.L.O.L. and M.L.P.v.d.H. contributed to the design of the study. A.Y., J.V., K.B. and M.D contributed to the data collection and extraction. A.Y., M.D. and N.v.G. performed the statistical analysis. A.Y., E.E.L.O.L., M.H. and N.v.G. interpreted the data. A.Y., M.D., M.L.P.v.d.H. and E.E.L.O.L. wrote the manuscript. All authors, including J.v.L., contributed to the discussion and revision of the manuscript. All authors have approved the final version.

Funding

No funding was received.

Conflict of interest

None to declare.

References

Barents EB, Bilo HJG, Bouma M, Van den Brink-Muinen A, Dankers M, Van den Donk M, Hart HE, Houweling ST, Ijzerman RG, Janssen PGH *et al*. Diabetes mellitus type 2. 2021. Nederlands Huisartsen Genootschap. https://richtlijnen.nhg.org/files/pdf/63_Diabetes%20mellitus%20type%202_november-2021.pdf.

Bashiri A, Giliutin M, Ziedenberg H, Plakht Y, Baumfeld Y. A proposed prognostic prediction tool for a live birth among women with recurrent pregnancy loss. *J Matern Fetal Neonatal Med* 2020; 1–7. [Online ahead of print]

Bender Atik R, Christiansen OB, Elson J, Kolte AM, Lewis S, Middeldorp S, Nelen W, Peramo B, Quenby S, Vermeulen N, *et al.*; ESHRE Guideline Group on RPL. ESHRE guideline: recurrent pregnancy loss. *Hum Reprod Open* 2018;**2018**:hoy004.

Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, Moons KG. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;**56**: 826–832.

Brigham SA, Conlon C, Farquharson RG. A longitudinal study of pregnancy outcome following idiopathic recurrent miscarriage. *Hum Reprod* 1999;**14**:2868–2871.

Cauchi MN, Coulam CB, Cowchock S, Ho HN, Gatenby P, Johnson PM, Lubs ML, McIntyre JA, Ramsden GH, Smith JB. Predictive factors in recurrent spontaneous aborters—a multicenter study. *Am J Reprod Immunol* 1995;**33**:165–170.

Cauchi MN, Pepperell R, Kloss M, Lim D. Predictors of pregnancy success in repeated miscarriage. *Am J Reprod Immunol* 1991;**26**: 72–75.

CBS. *Women Continue to Postpone Motherhood*. 2018.

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Ann Intern Med* 2015;**162**:735–736.

Copas JB. Regression, prediction and shrinkage. *J R Stat Soc Ser B (Methodol)* 1983;**45**:311–335. [CVOCROSSCVO]

Franssen MT, Korevaar JC, Leschot NJ, Bossuyt PM, Knegt AC, Gerssen-Schoorl KB, Wouters CH, Hansson KB, Hochstenbach R, Madan K *et al*. Selective chromosome analysis in couples with two or more miscarriages: case-control study. *BMJ* 2005;**331**:137–141.

Jaslow CR, Carney JL, Kutteh WH. Diagnostic factors identified in 1020 women with two versus three or more recurrent pregnancy losses. *Fertil Steril* 2010;**93**:1234–1243.

Jauniaux E, Farquharson RG, Christiansen OB, Exalto N. Evidence-based guidelines for the investigation and medical treatment of recurrent miscarriage. *Hum Reprod* 2006;**21**:2216–2222.

Liddell HS, Pattison NS, Zanderigo A. Recurrent miscarriage–outcome after supportive care in early pregnancy. *Aust N Z J Obstet Gynaecol* 1991;**31**:320–322.

Lund M, Kamper-Jorgensen M, Nielsen HS, Lidegaard O, Andersen AM, Christiansen OB. Prognosis for live birth in women with recurrent miscarriage: what is the best measure of success? *Obstet Gynecol* 2012;**119**:37–43.

McQueen DB, Zhang J, Robins JC. Sperm DNA fragmentation and recurrent pregnancy loss: a systematic review and meta-analysis. *Fertil Steril* 2019;**112**:54–60.e3.

Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012a; **98**:691–698.

Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, Grobbee DE. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012b;**98**:683–690.

NVOG. Herhaalde Miskraam *NVOG*. 2007. https://www.nvog.nl/wp-content/uploads/2017/12/Herhaalde-miskraam-2.0-08-06-2007.pdf.

Nybo Andersen AM, Hansen KD, Andersen PK, Davey Smith G. Advanced paternal age and risk of fetal death: a cohort study. *Am J Epidemiol* 2004;**160**:1214–1222.

Practice Committee of the American Society for Reproductive Medicine. Evaluation and treatment of recurrent pregnancy loss: a committee opinion. *Fertil Steril* 2012;**98**:1103–1111. [CVOCROSSCVO]

Quenby SM, Farquharson RG. Predicting recurring miscarriage: what is important? *Obstet Gynecol* 1993;**82**:132–138.

Rai R, Regan L. Recurrent miscarriage. *Lancet* 2006;**368**:601–611.

RCOG. *The Investigation and Treatment of Couples with Recurrent First trimester and Second trimester Miscarriage*. 2011. https://www.rcog.org.uk/globalassets/documents/guidelines/gtg_17.pdf.

Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, Moons KGM, Collins G, van Smeden M. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; **368**:m441.

Smith PP, Dhillon-Smith RK, O'Toole E, Cooper N, Coomarasamy A, Clark TJ. Outcomes in prevention and management of miscarriage trials: a systematic review. *BJOG* 2019;**126**:176–189.

Stephenson MD. Frequency of factors associated with habitual abortion in 197 couples. *Fertil Steril* 1996;**66**:24–29.

Sugiura-Ogasawara M, Ozaki Y, Kitaori T, Suzumori N, Obayashi S, Suzuki S. Live birth rate according to maternal age and previous number of recurrent miscarriages. *Am J Reprod Immunol* 2009;**62**: 314–319.

Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;**17**:230.

Venners SA, Wang X, Chen C, Wang L, Chen D, Guang W, Huang A, Ryan L, O'Connor J, Lasley B *et al.* Paternal smoking and pregnancy loss: a prospective study using a biomarker of pregnancy. *Am J Epidemiol* 2004;**159**:993–1001.

Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW, Koffijberg H, Moons KGM, Steyerberg EW. A closed testing procedure to select an appropriate method for updating prediction models. *Stat Med* 2017;**36**:4529–4539.

Youssef A, Lashley L, Dieben S, Verburg H, van der Hoorn ML. Defining recurrent pregnancy loss: associated factors and prognosis in couples with two versus three or more pregnancy losses. *Reprod Biomed Online* 2020;**41**:679–685.

Youssef A, Vermeulen N, Lashley E, Goddijn M, van der Hoorn MLP. Comparison and appraisal of (inter)national recurrent pregnancy loss guidelines. *Reprod Biomed Online* 2019;**39**:497–503.