



Universiteit
Leiden
The Netherlands

Identify, appraise and individualize: clinical practice and prediction models in recurrent pregnancy loss

Youssef, A.

Citation

Youssef, A. (2023, October 10). *Identify, appraise and individualize: clinical practice and prediction models in recurrent pregnancy loss*. Retrieved from <https://hdl.handle.net/1887/3643184>

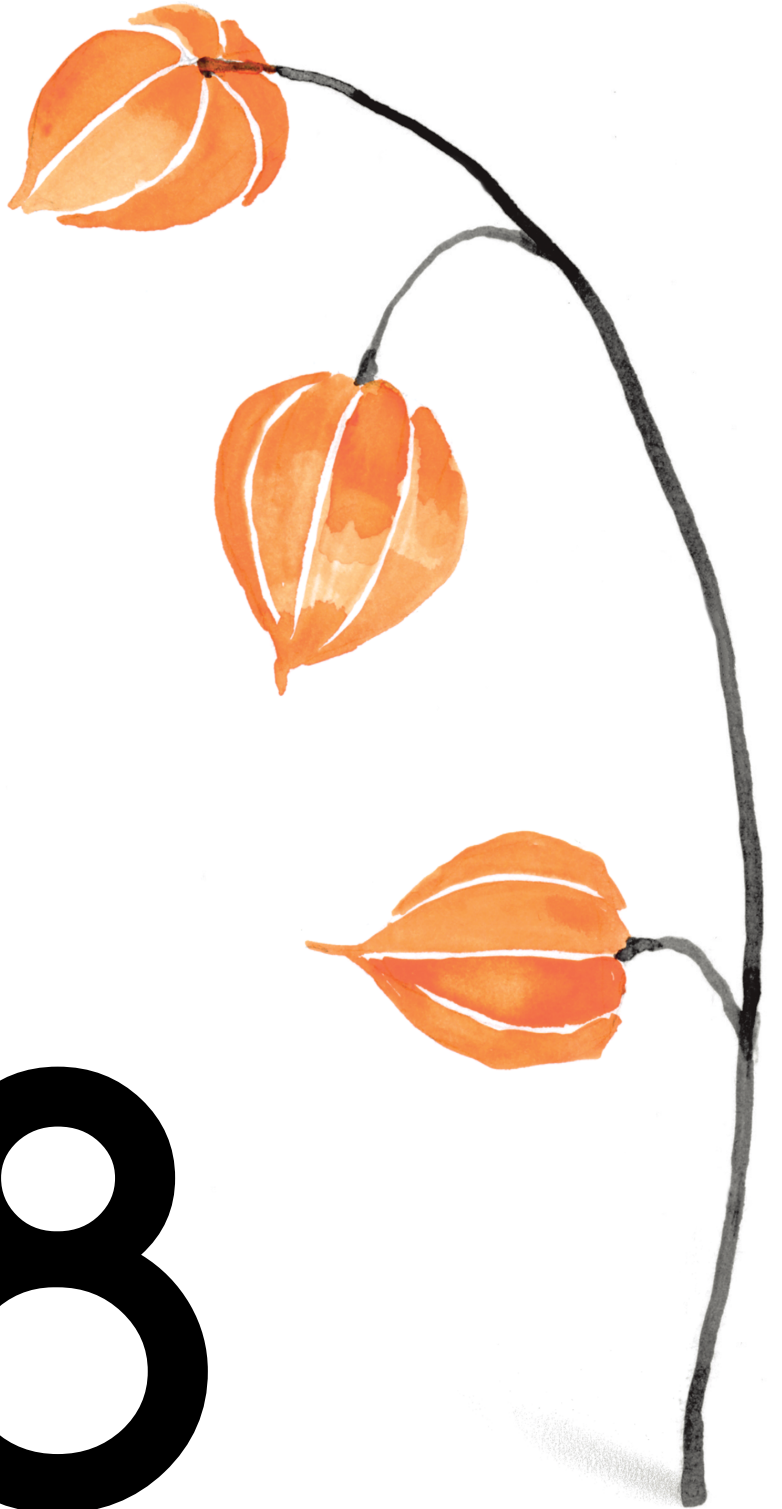
Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3643184>

Note: To cite this publication please use the final published version (if applicable).

8



CHAPTER 8

EXTERNAL VALIDATION OF A FREQUENTLY USED PREDICTION MODEL FOR ONGOING PREGNANCY IN COUPLES WITH UNEXPLAINED RECURRENT PREGNANCY LOSS

A. Youssef, M.L.P. van der Hoorn, M. Dongen, J. Visser, K.W.M.
Bloemenkamp, J.M.M. van N. van Geloven and E.E.L.O. Lashley

Human Reproduction, 2022

ABSTRACT

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 (LUMC) between 2004 and 2019.

PARTICIPANTS/MATERIALS, SETTING, METHODS

Unexplained RPL was defined as the loss of two or more pregnancies before 24 weeks, without presence of an identifiable cause for the pregnancy losses, according to the European Society for Human Reproduction and Embryology (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% (CI 95% 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 (CI 95% -0.62 - -0.31) and a calibration slope of 0.42 (CI 95% 0.11 - 0.73). The discriminative ability of the model was very low with a concordance statistic of 0.55 (CI 95% 0.51 - 0.59). Recalibration of the Brigham model hardly improved the c-statistic (0.57; CI 95% 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 in order to estimate the chance of a successful pregnancy in couples with unexplained RPL more accurately.

INTRODUCTION

Recurrent pregnancy loss (RPL) is defined as the loss of two or more conceptions (1). This condition affects 1-3% of all fertile couples (2, 3). RPL is a highly heterogeneous condition with multiple known maternal and paternal risk factors (4-6). Despite extensive diagnostic work-ups offered to couples with RPL, an underlying risk factor may be identified in only 25-50% of couples (7, 8). 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 is assumed to be of influence in the prevention of new pregnancy loss (9). 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 (10-16) and various international guidelines recommend the use of different prediction models (17). The European Society of Human Reproduction and Embryology (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 (1). The Brigham model has been implemented in RPL care in the Netherlands and in the United Kingdom (18, 19), while the American Society for Reproductive Medicine (ASRM) adapted the Lund model in their RPL guideline (20). 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 (21). 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.

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 presence of an identifiable cause for the pregnancy losses, according to the ESHRE guideline (1). The following investigations were performed to rule out factors associated with RPL: maternal testing for antiphospholipid syndrome (lupus antibodies, anticardiolipin antibodies, anti- β 2-glycoprotein antibodies), parental karyotyping for chromosomal abnormalities based on a priori chance (22), endocrinological factors (thyroid function and thyroid peroxidase antibody testing, random glucose level on indication (23)), and assessment of uterine cavity to rule out anatomic abnormalities. Testing for inherited thrombophilia and hyperhomocysteinemia was performed until 2018 as these were regarded 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 (10).

$$\log(p/(1-p))=2.00-0.0828 (\text{age}-32)-0.2467 (\text{number of pregnancy losses})$$

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 (24).

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. (25). 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. for the calculation of a sample size in clinical prediction models (26). 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% (27). 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 Leiden University Medical Centre (protocols P11.196 and P19.014).

RESULTS

Between 2004 and 2019, 904 couples with unexplained RPL were registered at the LUMC RPL clinic. 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 (Figure 1).

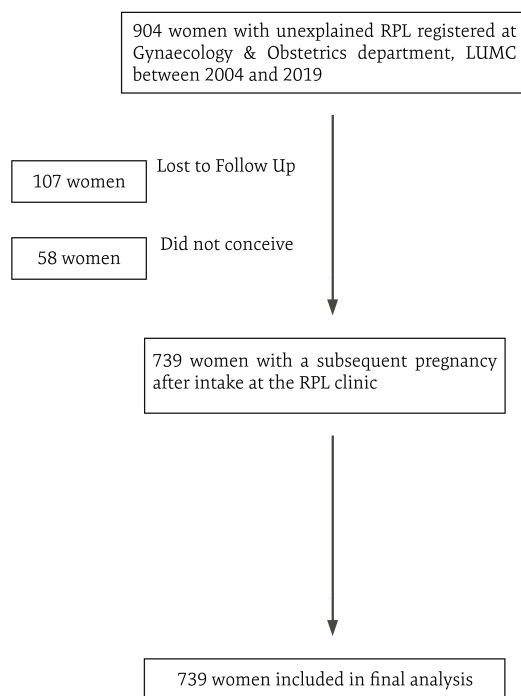


Figure 1. Flow chart of women with unexplained recurrent pregnancy loss who were considered for inclusion in external the validation.

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 shown in Table 1. 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% (CI 95% 60.6% - 67.6%) of couples, defined as a heartbeat on ultrasound ≥ 24 weeks pregnancy. Data of first pregnancy after intake is shown in Table 2.

Table 1 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	3941 (53.3%)
4	50 (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 2 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)

We plotted the expected success probabilities of the first pregnancy after intake according to Brigham's formula against the observed rates (Figure 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% (CI 95% for the 9.8% difference 6.3% - 13.3%)).

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

$\log(p/(1-p))=1.53-0.01(\text{age}-32)-0.28(\text{number of pregnancy losses})$

However, the updated model barely improved its discriminative ability (c-statistic 0.57; CI 95% 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 outcome of next pregnancy in couples with unexplained RPL, as developed by Brigham et al. (10). 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 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 (28). 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 suggests 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 (29). 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, that 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 (30). 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 (31, 32). In our study, we included patients from another geographical area and from another time period. This has influence on differences between the populations. First, the definition of RPL has significantly changed over the past 20 years (1). 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 recurrent pregnancy losses. 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, there is a need for standardised and patient-central clinical outcomes in studies on pregnancy after RPL (33).

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 tenth 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 in order to estimate the chance of a successful pregnancy in couples with unexplained RPL more accurately.

REFERENCES

1. The ESHRE Guideline Group on RPL, Bender Atik R, Christiansen OB, Elson J, Kolte AM, Lewis S, et al. ESHRE guideline: recurrent pregnancy loss. *Hum Reprod Open*. 2018;2018(2):hoy004.
2. Rai R, Regan L. Recurrent miscarriage. *Lancet*. 2006;368(9535):601-11.
3. Jauniaux E, Farquharson RG, Christiansen OB, Exalto N. Evidence-based guidelines for the investigation and medical treatment of recurrent miscarriage. *Human reproduction (Oxford, England)*. 2006;21(9):2216-22.
4. Nybo Andersen AM, Hansen KD, Andersen PK, Davey Smith G. Advanced paternal age and risk of fetal death: a cohort study. *American journal of epidemiology*. 2004;160(12):1214-22.
5. Venners SA, Wang X, Chen C, Wang L, Chen D, Guang W, et al. Paternal smoking and pregnancy loss: a prospective study using a biomarker of pregnancy. *American journal of epidemiology*. 2004;159(10):993-1001.
6. McQueen DB, Zhang J, Robins JC. Sperm DNA fragmentation and recurrent pregnancy loss: a systematic review and meta-analysis. *Fertility and sterility*. 2019.
7. Stephenson MD. Frequency of factors associated with habitual abortion in 197 couples. *Fertility and sterility*. 1996;66(1):24-9.
8. Jaslow CR, Carney JL, Kutteh WH. Diagnostic factors identified in 1020 women with two versus three or more recurrent pregnancy losses. *Fertility and sterility*. 2010;93(4):1234-43.
9. Liddell HS, Pattison NS, Zanderigo A. Recurrent miscarriage--outcome after supportive care in early pregnancy. *The Australian & New Zealand journal of obstetrics & gynaecology*. 1991;31(4):320-2.
10. Brigham SA, Conlon C, Farquharson RG. A longitudinal study of pregnancy outcome following idiopathic recurrent miscarriage. *Human reproduction (Oxford, England)*. 1999;14(11):2868-71.
11. Cauchi MN, Coulam CB, Cowchock S, Ho HN, Gatenby P, Johnson PM, et al. Predictive factors in recurrent spontaneous aborters--a multicenter study. *American journal of reproductive immunology (New York, NY : 1989)*. 1995;33(2):165-70.
12. Cauchi MN, Pepperell R, Kloss M, Lim D. Predictors of pregnancy success in repeated miscarriage. *American journal of reproductive immunology (New York, NY : 1989)*. 1991;26(2):72-5.
13. 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? *Obstetrics and gynecology*. 2012;119(1):37-43.
14. Quenby SM, Farquharson RG. Predicting recurring miscarriage: what is important? *Obstetrics and gynecology*. 1993;82(1):132-8.

15. 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. *American Journal of Reproductive Immunology*. 2009;62(5):314-9.
16. 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.
17. Youssef A, Vermeulen N, Lashley E, Goddijn M, van der Hoorn MLP. Comparison and appraisal of (inter)national recurrent pregnancy loss guidelines. *Reproductive biomedicine online*. 2019.
18. NVOG. Herhaalde Miskraam 2007 [08/06/2007:[Available from: www.nvog-documenten.nl/richtlijn/doc/download.php?id=750].
19. RCOG. The Investigation and Treatment of Couples with Recurrent Firsttrimester and Second-trimester Miscarriage. 2011.
20. Practice Committee of the American Society for Reproductive M. Evaluation and treatment of recurrent pregnancy loss: a committee opinion. *Fertility and sterility*. 2012;98(5):1103-11.
21. 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(10):735-6.
22. Franssen MT, Korevaar JC, Leschot NJ, Bossuyt PM, Kneeght AC, Gerssen-Schoorl KB, et al. Selective chromosome analysis in couples with two or more miscarriages: case-control study. *BMJ (Clinical research ed)*. 2005;331(7509):137-41.
23. Barents EB, HJG; Bouma, M; Van den Brink-Muinen, A; Dankers, M; Van den Donk, M; Hart, HE; Houweling, ST; IJzerman, RG; Janssen, PGH; Kerksen, A; Palmen, J; Verburg-Oorthuizen, AFE; Wiersma, T. Diabetes mellitus type 2: Nederlands Huisartsen Genootschap; 2018 [Available from: <https://richtlijnen.nhg.org/standaarden/diabetes-mellitus-type-2>].
24. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic t, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230.
25. Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW, et al. A closed testing procedure to select an appropriate method for updating prediction models. *Stat Med*. 2017;36(28):4529-39.
26. Riley RD, Ensor J, Snell KIE, Harrell FE, Jr., Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ (Clinical research ed)*. 2020;368:m441.
27. 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. *Reproductive biomedicine online*. 2020;41(4):679-85.

28. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol.* 2003;56(9):826-32.
29. CBS. Women continue to postpone motherhood 2018 [Available from: <https://www.cbs.nl/en-gb/news/2018/05/women-continue-to-postpone-motherhood>].
30. Copas JB. Regression, Prediction and Shrinkage. *Journal of the Royal Statistical Society: Series B (Methodological).* 1983;45(3):311-35.
31. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart.* 2012;98(9):691-8.
32. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart.* 2012;98(9):683-90.
33. 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(2):176-8

