Research Article

External Validation of Models to Predict Unsuccessful Endometrial Ablation: A Retrospective Study

Stevens Kelly Yvonne Roger[1,2*], Muller Iris[4], Houterman Saskia[3], Weyers Steven[2], Schoot Benedictus[1,2]

[1]Department of Obstetrics and Gynaecology, Catharina Hospital, Eindhoven, the Netherlands, Michelangelolaan 2, 5623 EJ Eindhoven, the Netherlands

[2]Women's Clinic, Ghent University Hospital, Ghent, Belgium, Corneel Heymanslaan 10, 9000 Ghent, Belgium

[3]Department of Education and Research, Catharina Hospital, Eindhoven, the Netherlands, Michelangelolaan 2, 5623 EJ Eindhoven, the Netherlands

[4]Department of Obstetrics and Gynaecology, ZGT, Almelo, the Netherlands, Zilvermeeuw 1, 7609 PP Almelo, the Netherlands

[*]**Corresponding Author:** Stevens Kelly Yvonne Roger, Department of Obstetrics and Gynaecology, Catharina Hospital, Eindhoven, the Netherlands, Michelangelolaan 2, 5623 EJ Eindhoven, the Netherlands

Abstract

Study Objective

External validation of our previously presented and locally established prediction models to help counsel patients for failure of endometrial ablation (EA) or surgical re-intervention within 2 years after EA, called 'Failure model' and 'Re-intervention model' respectively.

Design

Retrospective external validation study, minimal follow-up time of 2 years.

Setting

Two non-academic teaching hospitals in the Netherlands.

Patients

Pre-menopausal women (18+) who had undergone EA

for abnormal uterine bleeding problems between January 2010 and November 2012. A total of 329 patients were eligible for analysis.

Interventions

Interventions used for EA were Novasure (Hologic, Marlborough, Massachusetts, US) and ThermaChoice III (Ethicon, Sommerville, US).

Measurements and Main Results

The Area Under the Receiver Operating characteristics Curve (AUROC) for the outcome parameter of failure within 2 years after EA was 0.59 (95% CI 0.53 – 0.65). Variables in this model were dysmenorrhea, age, parity ≥5 and preoperative menorrhagia. The Hosmer-Lemeshow test showed no significant difference between the observed and predicted outcome. (Chi-square: 4.62, P-value: .80) The AUROC for the outcome parameter surgical re-intervention within 2 years was 0.62 (95% CI 0.53 – 0.70) Variables in this model were dysmenorrhea, age, menstrual duration>7 days, parity ≥5 and a previous caesarean section. The Hosmer-Lemeshow test showed no significant difference between the observed and predicted outcome (Chi-square 11.34, P-value .18).

Conclusion

Both the failure model and the re-intervention model can be used to predict unsuccessful endometrial ablation in the general population within two years after the procedure. It can be used prior to surgery to facilitate tailor-made shared decision-making, and help counsel patients with regards to the potential outcome of their treatment with the use of a personally calculated percentage.

Keywords: Endometrial ablation; External validation; Prediction model

1. Introduction

Endometrial ablation (EA) is frequently used as treatment for the common gynaecologic problem: heavy menstrual bleeding. It is increasingly used because of its minimally invasive character, low costs, low risks and short recovery time [1-4]. In 2017, approximately 9000 endometrial ablations were performed in the Netherlands, whereas in the US it was stated to be around 400.000 procedures [5]. Short term success-rates up to the period of one year have suggested that EA is highly effective, however, long-term follow-up shows diminishing results. In fact, prevalence of post- EA hysterectomy can be as high as 20%, mainly due to complaints of pain or abnormal uterine bleeding [6-8]. Current literature is inconclusive about which variables influence the outcomes of EA. For this reason, we previously developed two internally validated prediction models [9]. The first model, called the *'Failure model'*, showed variables significant in predicting EA failure. Failure was defined as: patient dissatisfaction, lower abdominal pain or complaints of abnormal uterine bleeding after EA. Significant variables were age, dysmenorrhea, parity ≥5 and preoperative menorrhagia. The AUC after internal validation was 0.68 [9]. The second model called the *'Re-intervention model'*, predicted the outcome of surgical re-intervention within 2 years after EA. Significant variables were age, dysmenorrhea, menstrual duration> 7 days, parity ≥5 and previous caesarean section. The AUC after internal validation was 0.71 [9]. These internally validated models can be used to help counsel patients with regards to the potential outcome of their treatment with the use of a personally calculated percentage. In order to encourage a wider use of these models, the aim of this study is to externally validate both models, so that they can be implemented for patient counselling in the general population.

Objective

The objective of this study was to construct an external validation of the previously published internally validated 'Failure Model' and 'Re-intervention model' by Stevens et al. [9] with the use of an external dataset.

2. Methods

Study design

This retrospective external validation study used data from Ziekenhuisgroep Twente (ZGT) Almelo/Hengelo' and ' Medisch Spectrum Twente (MST) Enschede', both non-academic teaching hospitals in the Netherlands. This external dataset was used in the study from Muller et al. [10] which measured patients satisfaction and amenorrhea rate after EA, and included patients who had undergone EA between January 2010 and November 2012. Similar ablation techniques were used in both hospitals, namely: ThermaChoice III ® (Ethicon, Sommerville, US) and Novasure ® (Hologic, Marlborough, Massachusetts, US). Previous literature showed that these techniques are equally effective [11-13]. The full study protocol has been previously published, all patients included in the study of Muller et al. gave their informed consent [10].

Patients

This external dataset included pre-menopausal women (18+) undergoing endometrial ablation due to abnormal uterine bleeding. Women who had a (suspicion of) malignancy or cavity-deforming abnormalities were excluded for external validation. Furthermore, women were excluded if the endometrial ablation could not be, or was incompletely performed. The duration of follow-up was at least two years after EA because, as stated in our previous article [9], literature has shown that most re-interventions take place within this time frame [9,13,14]. Follow-up ended on the day of hysterectomy, in case of death, or on 30th of March 2017 after the last chart review was done. Similar criteria for patient selection (in the internal dataset) were used in the article of Stevens et al. were the internally validated prediction models were made. The prediction model as validated in this article was constructed based in an (internal) dataset of patient outcomes in our hospital [9]. The term ' External dataset' refers to the patient outcomes from a study in a regional hospital in the eastern part of our country as published by Muller et al. [10]. This specific external dataset will be used to validate our prediction models in the present study. The study was performed in accordance with the relevant guidelines and regulations.

Data extraction

The external dataset from Muller et al. provided us with the majority of the required information [10]. Extra patient chart review was done by two of our researchers to collect additional relevant data (for example pathology results) where necessary. Data regarding one significant factor in the previously published re-intervention model 'duration of menstruation > 7 days' could not be obtained. It was unfortunately neither described in the given dataset, nor in electronic patient records.

Statistical analyses

The baseline characteristics of the patients in the internal and external dataset were compared. Categorical variables were reported as numbers and frequencies, and continuous variables as means with standard deviations or median and minimum-maximum, depending on normality. Between group differences were assessed by the independent t-test for continuous variables if they were normally distributed,

and the Mann- Whitney U test if not. The Chi-square test was used to compare categorical variables between groups. A P-value < .05 was considered significant. The predicted probability of both models (P-Failure or P-Re-intervention) was calculated by using the previously made internally validated prediction models for failure of EA and surgical re-intervention respectively [9]. The internally validated formulas for the calculated probability were as follows:

$$\textit{P-Failure} = 1/(1+EXP(-Y1)).$$

Y1= 3.485 - (age * 0.063) + (dysmenorrhea (yes = 1, no = 0) * 0.677) + (parity ≥ 5 (yes = 1, no = 0) * 2.183) - (menorrhagia (yes = 1, no = 0) * 1.400).

$$\textit{P- Re-intervention} = 1/(1+EXP(-Y2)).$$

Y2= -0.896 - (age * 0.046) + (duration of menstruation >7 days (yes = 1, no = 0) * 0.629 + (dysmenorrhea (yes = 1, no = 0) * 0.00794) + (parity ≥ 5 (yes = 1, no = 0) * 1.781) + (previous caesarean section (yes = 1, no = 0) * 0.700).

Area Under the Receiver Operating characteristics Curve (AUROC) and Nagelkerke's R square were used to evaluate model performance [15,16]. The AUROC was used to test the discriminative value of the models. AUROC ranges from 0.0 to 1.0. where a value of 0.5 indicates that a model does not predict an outcome better than random chance. Therefore, 0.5 should be considered as the minimum value of AUROC [15-18]. Calibration of the models was tested by using the Hosmer-Lemeshow goodness-of-fit test and calibration plots. This assesses the hypothesis of the perfect agreement between the predicted and observed outcomes [15-17]. The slope and intercept of the regression line in the calibration plot were calculated for both models. A slope of one and an intercept of

zero indicates a perfect calibration, were the predicted and the observed outcomes are a perfect fit [16,19]. In the re-intervention model, 'duration of menstruation > 7 days' was a relevant factor [9]. However, in the external dataset, this variable was not available [10]. Therefore, we performed a sensitivity analysis by calculating the discrimination and calibration in three different manners. In this way we could evaluate the necessity of including this variable into the analysis. This was tested as follows: first, we tested the model with the 'duration of menstruation <7 days' (DMno) in all the cases, following by testing with a 'duration of menstruation >7 days' present in all the cases (DMyes). Next, it was repeated with 40% of the cases having a duration of menstruation >7 days (DM40). We choose 40% incidence (DM40), because this was comparable to the incidence in the internal validation group [9]. IBM SPSS statistics, software version 26.0 (IBM Corp., Armonk, NY, USA) was used to perform statistical analysis. The TRIPOD guidelines for validation studies were taken into account when performing the external validation [16].

3. Results

In the external dataset, a total of 613 patients were selected (345 Novasure and 268 ThermaChoice III). 98 patients were excluded because of current postmenopausal status, and one patient deceased. Another seven patients were excluded from analysis because of a cavity deforming abnormality (e.g. myoma). Additionally, 23 patients were excluded because of a failed procedure. Of the remaining 484, 155 patients did not fill in the questionnaire in the study from Muller et al. [10]. In total, 329 patients were available for analysis, giving a response-rate of 68% (Figure 1). Baseline characteristics of both the internal and external dataset are summarized in table 1. In the external dataset, median age was 43 years (range

30 - 51 years), median BMI was 25.6 kg/m2 (17.7 – 46.1 kg/m^2) and median parity was 2 (range 0 – 9). Dysmenorrhea was found in 79.3% of patients, menorrhagia in 98.1%, and 21.9% had a previous caesarean section. In the internal dataset, median age was 44 years (range 25 - 55), median BMI was 25.5 kg/m^2 (range 18.3 – 46.6) and median number of parity was 2 (range 0-6). Dysmenorrhea was found in 57.4% of patients, menorrhagia in 97.3%, a total of 13.7% have had a previous caesarean section. Due to the inclusion criterium of the previous performed endometrial ablations, the prevalence of patients with menorrhagia in both the internal and external database is high. A significant difference was found between the group of the internal- and external dataset on the following points: age (P-value .001), previous caesarean section (P-value .003) and dysmenorrhea (P-value <.001). No further differences in baseline characteristics were found between the groups.
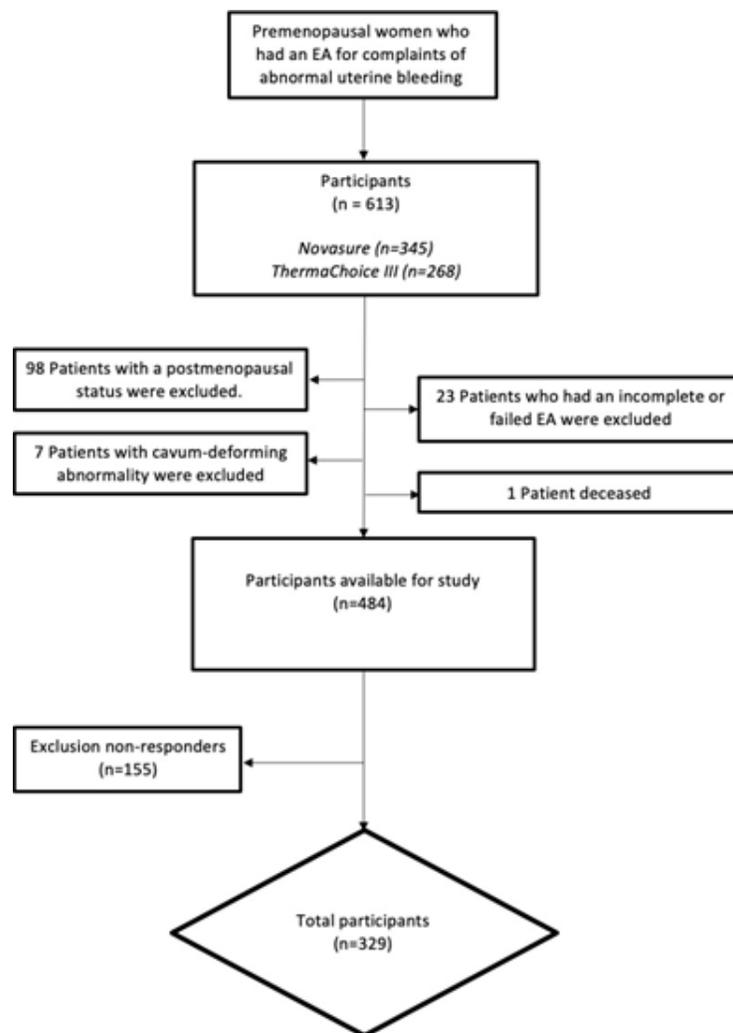

Figure 1: Enrolment and allocation of patients who have had an endometrial ablation for complaints of heavy menstrual bleeding

In both the internal and external dataset, 11.9% had a surgical re-intervention within two years after the ablation (P-value .99). There was also no significant difference between the groups for the outcome measure of treatment failure, being 35.8% and 34.1% respectively (P-value .64). Further results showed no difference in hysterectomy rate in the external dataset (18.5%) versus 18.8% in the internal dataset (P-value .92). Of the 61 patients in the external dataset who had a hysterectomy, 49% (n=30) of the histopathology results showed adenomyosis. Two patients had signs of endometriosis, and nine patients were diagnosed with a uterine myoma. Seven of the 30 patients with adenomyosis had concomitant myomas (table 1 here).

Characteristic	Internal validation dataset (N=446)		External validation Dataset (N=329)		P-Value
	N	Frequency, mean or median*	N	Frequency, mean or median*	
Age (y)	446	44 (25-55)	329	43 (30-51)	0.001
Body mass index (kg/m²)	446	25.5 (18.3-46.6)	302	25.6 (17.7 –46.1)	0.465
Dysmenorrhea	446	57.40%	323	79.30%	< .001
Duration of menstruation >7 days	429	39.40%		Not reported	
Length of the uterus (cm) Menorrhagia	402	9 (5-14)	306	9 (6-12)	0.711
Menorrhagia	446	97.30%	324	98.10%	0.447
Parity (No.)	446	2 (0-6)	329	2 (0-9)	0.501
Previous caesarean section	446	13.70%	329	21.90%	0.003
Smoking	445	21.60%	211	24.60%	0.379
Sterilization	444	26.10%	328	23.20%	0.348

*Categorical variables are mentioned as frequencies, continuous variables as mean and standard deviation or median and minimum-maximum, depending on normality.

Table 1: Baseline Patient Characteristics of pre-menopausal women who had an EA, characteristics from both the internal and external dataset are mentioned.

External validation

Failure model
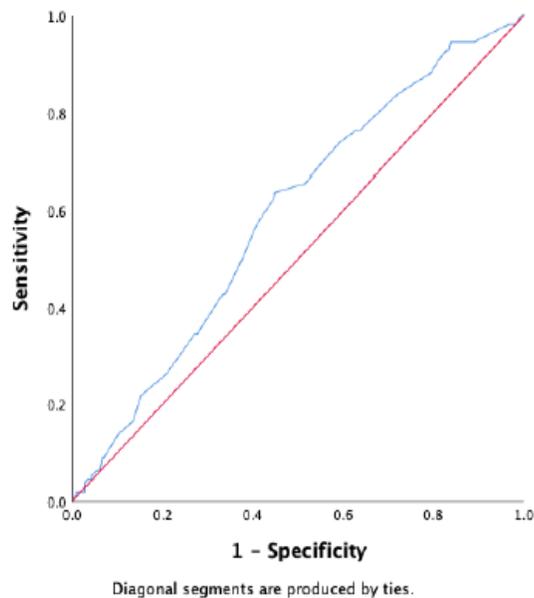

Diagonal segments are produced by ties.

Figure 2: ROC-curve external validation of the failure model. The diagonal is the reference line, indicating an AUC of 0.50, which indicates that a model predicts the same as random chance.

In the external dataset a total of 34% experienced ablation failure. When the failure model was used on the external dataset, it showed a moderate discriminative capacity, with an AUROC of 0.59 (95% CI 0.53 – 0.65) (Figure 2). Nagelkerke's R square for the overall performance of the failure model was 0.050. The Hosmer-Lemeshow test showed no significant difference between the observed and predicted failure (Chi-square: 4.62, P-value: .80). The predicted probability outcomes of the failure model are compared with the observed outcomes as seen in the calibration plot in figure 3. All the points (which represent our cases) are above the reference line, indicating that the predicted probabilities of failure are underestimated. The intercept of the calculated regression line is 0.19, which indicates that the predicted failure rates are too low. The estimated slope was 0.89, meaning that the probabilities are also too optimistic. Figure 4 shows the distribution of the predicted failure rates per patient by using the failure model. Most patients had a failure rate between 42% and 62%. A failure rate of $\geq$ 85% was seen in 6 patients. No patients had a failure rate above 94% or under 24%.
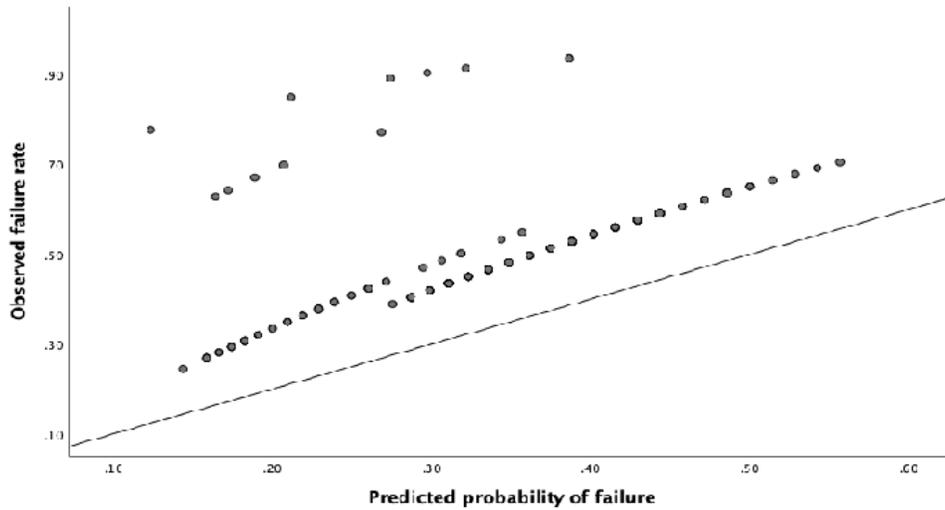
Figure 3: Calibration plot, showing relationship between observed and predicted failure rate. When the points in the plot are in exact line with the diagonal reference line, the model predicts perfectly, because there is a perfect agreement between the predicted and the observed failure rates.
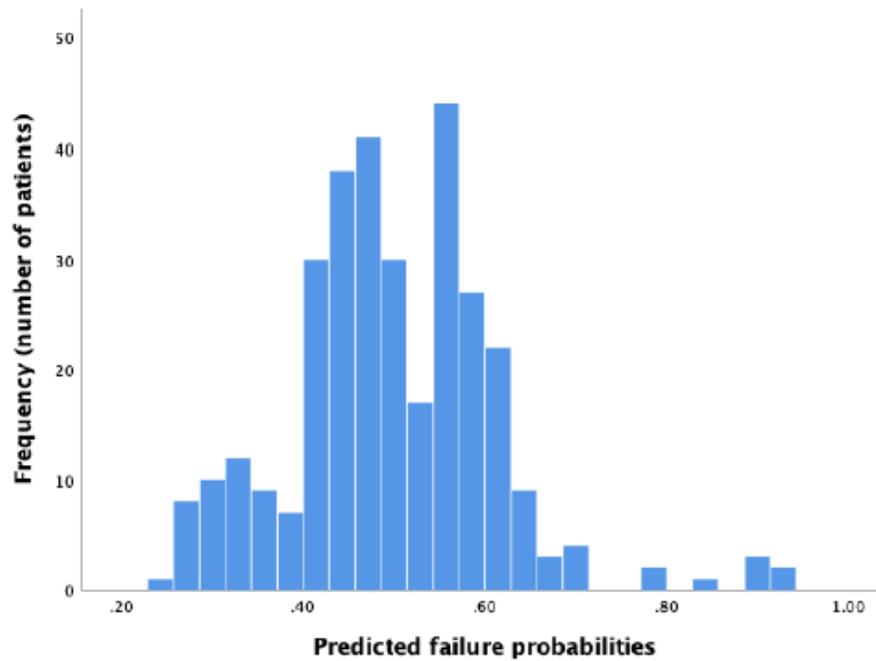

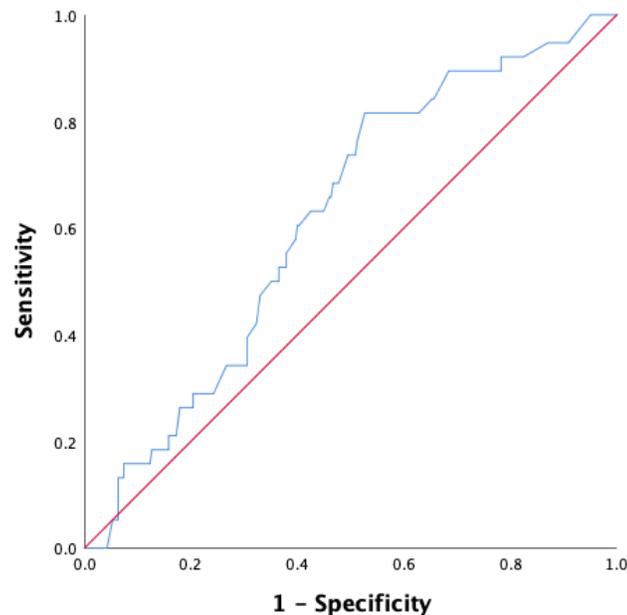**Figure 4:** Distribution of predicted failure rates using the failure model.

Re- intervention model

Figure 5: ROC-curve external validation of re-intervention model. The diagonal is the reference line, indicating an AUC of 0.50, which indicates that a model predicts the same as random chance

In the external dataset, 11.9% of women had a surgical re-intervention within two years after ablation. When the re-intervention model was used on the external dataset, it showed a moderate discriminative capacity, with an AUROC of 0.62 (95% CI 0.53 – 0.70) (Figure 5). Sensitivity analyses showed no significant difference when 'duration of menstruation > 7 days' was present in all the cases (DMyes), present in 40% of the cases (DM40) or in neither of the cases (DMno). The 95% CI of the different settings were overlapping. Nagelkerke's R square for the overall performance of the re-intervention model was 0.065. The Hosmer-Lemeshow test showed no significant difference between the observed and predicted outcome (Chi-square 11.34, P-value .18). The predicted probability outcomes of the re-intervention model are compared with the observed outcomes as seen in the calibration plot in figure 6. In our case, most of the points are beneath or in line with the reference line. The intercept of the calculated regression line is 0.04 which indicates that predicted re-intervention rates are too low. The estimated slope was 0.44 meaning that the probabilities are too optimistic, and are reflecting overfitting. Combining these outcomes, this means that in this case, as seen in figure 6, the high predicted re-intervention rates are too high and the low probabilities are too low. Figure 7 shows the distribution of the predicted re-intervention rates per patient by using the re-intervention model. Most patients had a re-intervention rate between 4% and 22%. No patients had a re- intervention rate above 58% or under 4%.
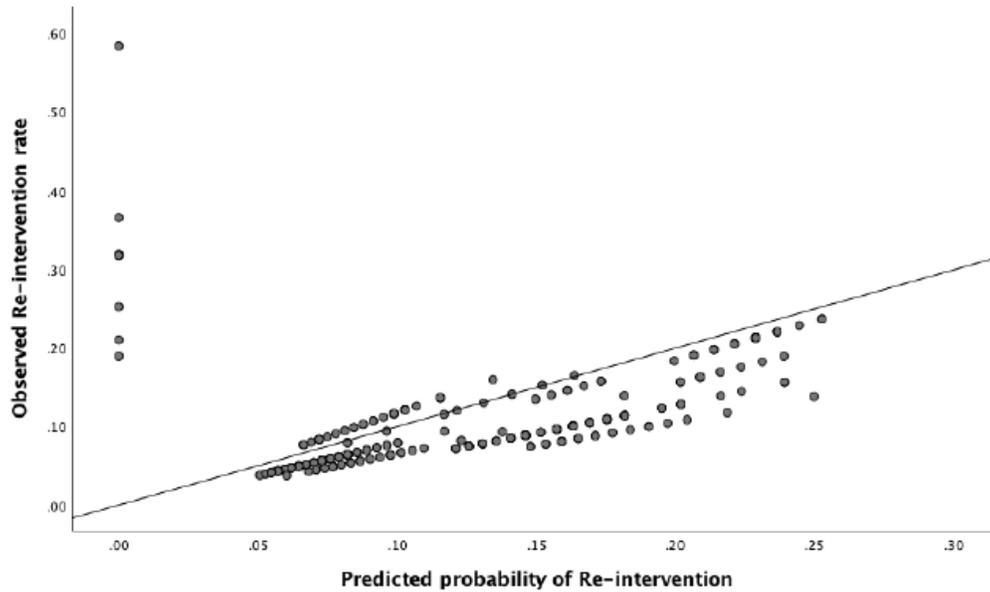
Figure 6: Calibration plot, showing relationship between observed and predicted Re-intervention rate. When the points in the plot are in exact line with the diagonal reference line, the model predicts perfectly, because there is a perfec agreement between the predicted and the observed failure rates.
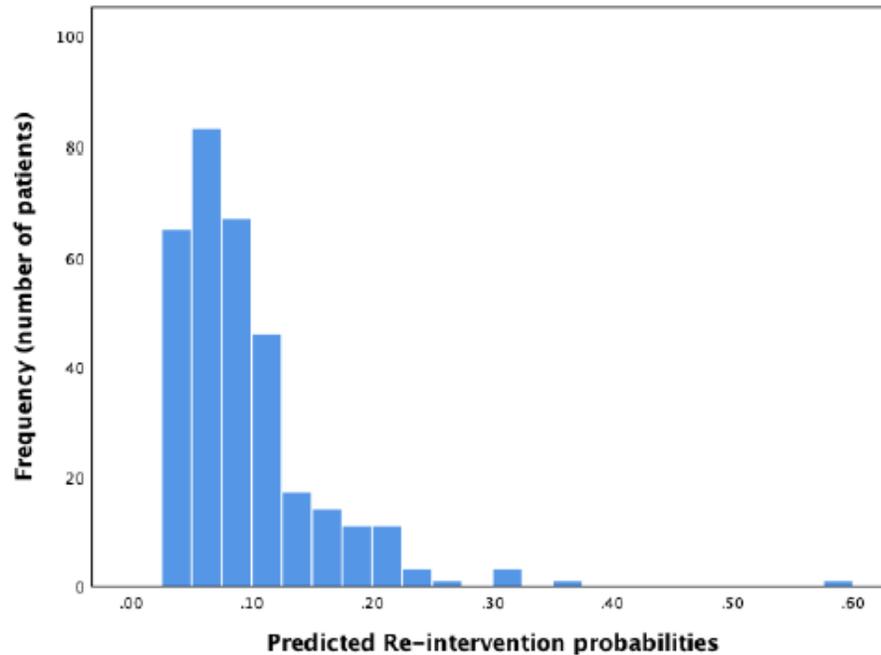

Figure 7: Distribution of predicted re-intervention rates using the re-intervention model.

4. Discussion

Since internal validation of our models to predict unsuccessful endometrial ablation was promising [9], external validation was performed to examine if the models could be widely applicable and useful for the general population. Explanation of the models' significant variables, consistent with literature, [7,20-22] can be found in the article of Stevens et al. [9]. We are aware of the fact that other variables may play a role [20-23]. Cavity-deforming abnormalities were excluded from the selection, some studies say however that intramural myomas can influence the outcome of EA as well [21]. However, other literature shows that only large submucosal myomas are a risk factor for failure of EA [22], therefore, this group was excluded from analysis. Furthermore, the number of myomas in this group was so small that they could not influence the outcome of our prediction model. El Nashar et al [20], also made a EA failure model, however, to the best of our knowledge, this model was not externally validated. At baseline, a significant difference in age was seen between the internal and external dataset. However, this difference is only one year, which does not seem clinically relevant. A second difference was seen in the patients with a previous caesarean section and patients with dysmenorrhea. Based on our internal validated models, both factors give higher chance on failure or re-intervention after EA. However, it is important that no significant difference was seen between the internal and external dataset in the model's outcome measures of failure or re-intervention. A possible explanation of the baseline difference in dysmenorrhea could be the subjective character of this variable. Moreover, 49% of hysterectomy pathology results in the external dataset showed signs of adenomyosis. This may explain the high level of dysmenorrhea in the external dataset. It is confusing that despite literature on adenomyosis as

factor for unsuccessful EA, it had no effect on the number of patients with dysmenorrhea included in the external dataset [7,20,22,24]. Patient selection is therefore important, suggesting patients with dysmenorrhea should be screened for adenomyosis, using for example the recently developed MUSA criteria [25]. The baseline difference in patients with previous caesarean section can possibly be found in the increasing interest in uterine scar defects (niches) and subsequent bleeding problems over the last years [26,27]. Patient-selection in the external dataset was between 2010-2012 [10], the internal between 2004-2013 [9]. Although there is a fairly uniform policy in the Netherlands with regard to treatment and diagnosis of abnormal uterine bleeding, it is possible that pathophysiology of the niche is approached differently in various parts of the country. In short, we are of the opinion that awareness for both dysmenorrhea and previous caesarean section in people wanting EA is important. After external validation, both the re-intervention model and failure model can be used in the general population with an moderate AUROC of 0.62 and 0.59 respectively. It should be noted that there still is a certain degree of inaccuracy. Although the results of the AUROC are moderate, these prediction models can provide the clinician a tool to discuss the pros and cons prior to surgery. Further research could focus on performing model updating [16].

Strengths and limitations

The fact that this study design requires retrospective data can be seen as a limitation. This design has a higher chance of missing data, as in our case, the variable 'duration of menstruation >7 days' of the re-intervention model was not known. However, we performed a sensitivity analysis, which showed no significant difference. Strengths were the multicentre

design and the extra chart review done by two different researchers. Since the participating hospitals for the external- and internal validation were in different parts of our (small) country, this validation can be seen as geographical validation. The models are made by logistical regression, however, it's also possible to use machine learning (ML). Hence, a study was conducted to see if ML can create better models than models made with logistic regression. This study showed that for the outcome re-intervention, logistic regression is a better predictor [28]. However, it is important to keep ML in consideration. Especially in large datasets with variables with strong predictive power [29,30] and small amount of pre-defined variables in the model [29,31-33]. Prediction models can be used for patient counselling, hoping that the uncertainness of patients can be assuaged with better insight into outcome of their treatment. We can optimise the shared decision making process, and allow patients to make a decision based on their personal calculated percentage. One notable issue with this approach however, is that the interpretation of percentages can be individual-specific. Some patients (or doctors) may find a failure rate of 30% acceptable, while for others this might be 75%. This encourages the conducting of research into not only the outcomes of prediction models, but how their results influence the (clinical) decision making of both patient and doctor.

Using the model

To facilitate general use of the models, a website was made:

https://www.prediction-failure-of-endometrialablation.com

Different patient characteristics can be filled in, and the individual calculated percentage of re-intervention and failure will be provided. These models can be used during consultations to support patient counselling.

5. Conclusion

After the performance of the external validation, both the re-intervention model and the failure model can be used to predict unsuccessful endometrial ablation within two years after the procedure. The outcome failure is defined as: complaints of abnormal uterine bleeding, patient dissatisfaction or lower abdominal pain. The outcome re-intervention is defined as any surgical re-intervention within 2 years after the EA. Both of these models, used prior to treatment, can facilitate patient counselling and support the tailor-made shared decision-making process regarding EA for the general population.

Disclosures

Availability of data and material

The datasets generated and analysed during the current study are not publicly available due to privacy, but they are available from the corresponding author on a reasonable request.

Competing interests

B.C. Schoot received fees from Medtronic on an hourly basis for lectures on hysteroscopic morcellation. The fees were donated to a foundation that promotes research in obstetrics and gynaecology. The remaining authors have no competing interests.

Consent for publication

Provided

Funding

None

IRB approval

This study was exempt from Regional Ethics Review Board approval, under the legal requirements for clinical research in the Netherlands.

Authors contribution

K.Y.R.Stevens: Project development, Data collection/management, Data analysis, Manuscript writing/editing

S. Houterman: Data analysis, Manuscript editing

S. Weyers: Manuscript editing

I. Muller: Data collection, Manuscript editing

B.C. Schoot: Project development, Manuscript editing

References

1. Karimi-Zarchi M, Fathi M, Tabatabaie A, et al. Long- term outcome of endometrial ablation therapy with cavaterm thermal balloon in patients with abnormal uterine bleeding. J Turkish Ger Gynecol Assoc. (2020).

2. Wortman M. Endometrial Ablation: Past, Present, and Future Part II. Surgical technology international. (2018).

3. Kuchenbaecker KB, McGuffog L, Barrowdale D, et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. J Natl Cancer Inst 109 (2017): 10-14.

4. Bergeron C, Laberge PY, Boutin A, et al. Endometrial ablation or resection versus levonorgestrel intra-uterine system for the treatment of women with heavy menstrual bleeding and a normal uterine cavity: A systematic review with meta-analysis. Human Reproduction Update (2020).

5. Charles E, Miller M. Diagnosis and treatment of global endometrial ablation failure. MDEdge (2016).

6. Longinotti MK, Jacobson GF, Hung YY, et al. Probability of hysterectomy after endometrial ablation. Obstet Gynecol 112 (2008): 1214-1220.

7. Wishall KM, Price J, Pereira N, et al. Postablation risk factors for pain and subsequent hysterectomy. Obstet Gynecol [Internet]. 124 (2014): 904-910.

8. Klebanoff J, Makai GE, Patel NR. Incidence and predictors of failed second-generation endometrial ablation. Gynecol Surg 14 (2017): 26.

9. Stevens KYR, Meulenbroeks D, Houterman S, Gijsen T, Weyers S, Schoot BC. Prediction of unsuccessful endometrial ablation: a retrospective study. Gynecol Surg 16 (2019): 7.

10. Muller I, Van der Palen J, Massop-Helmink D, et al. Patient satisfaction and amenorrhea rate after endometrial ablation by ThermaChoice III or NovaSure: a retrospective cohort study. Gynecol Surg 12 (2015): 81-87.

11. Louie M, Wright K, Siedhoff MT. The case against endometrial ablation for treatment of heavy menstrual bleeding. Curr Opin Obstet Gynecol 30 (2018): 287-292.

12. Lethaby A, Penninx J, Hickey M, et al. Endometrial resection and ablation techniques for heavy menstrual bleeding. Cochrane database Syst Rev 8 (2013): 15-23.

13. Bofill RM, Lethaby A, Grigore M, et al. Endometrial resection and ablation techniques for heavy menstrual bleeding. Cochrane Database of Systematic Reviews (2019).

14. Cooper K, Lee AJ, Chien P, et al. Outcomes following hysterectomy or endometrial ablation for heavy menstrual bleeding: Retrospective analysis of hospital episode statistics in Scotland. BJOG An Int J Obstet Gynaecol 8 (2011): 13-18.

15. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation and Updating (2019).

16. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med (2015).

17. Hosmer DW, Lemeshow S. Applied logistic regression. (2nd edtn), John Wiley & Sons, Inc (2000).

18. Hajian TK. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian Journal of Internal Medicine (2013).

19. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology (2010).

20. El-Nashar SA, Hopkins MR, Creedon DJ, et al. Prediction of treatment outcomes after global endometrial ablation. Obstet Gynecol 113 (2018): 97-106.

21. Lybol C, Van der Coelen S, Hamelink A, et al. Predictors of Long- Term NovaSure Endometrial Ablation Failure. J Minim Invasive Gynecol (2018).

22. Beelen P, Reinders IMA, Scheepers WFW, et al. Prognostic Factors for the Failure of Endometrial Ablation: A Systematic Review and Meta-analysis. Obstetrics and Gynecology (2019).

23. Moulder JK, Yunker A. Endometrial ablation: considerations and complications. Curr Opin Obstet Gynecol 28 (2016): 261-266.

24. Karpathiou G, Chauleur C, Dal Col P, et al. Histologic findings in hysterectomies after endometrial ablation. Pathol Res Pract (2020).

25. Van Den Bosch T, Dueholm M, Leone FPG, Valentin L, Rasmussen CK, Votino A, et al. Terms, definitions and measurements to describe sonographic features of myometrium and uterine masses: A consensus opinion from the Morphological Uterus Sonographic Assessment (MUSA) group. Ultrasound Obstet Gynecol (2015).

26. Morris H. Surgical pathology of the lower uterine segment caesarean section scar: Is the scar a source of clinical symptoms? Int J Gynecol Pathol (1995)

27. Tower AM, Frishman GN. Cesarean Scar Defects: An Underrecognized Cause of Abnormal Uterine Bleeding and Other Gynecologic Complications. Journal of Minimally Invasive Gynecology (2013).

28. Stevens KYR, Lagaert LVR, Bakkes T, et al. Clinical Prediction of Unsuccessful Endometrial Ablation: Random Forest vs Logistic Regression. J Minim Invasive Gynecol 26 (2019): 10-19.

29. Evangelia christodoulou, Jie MA, Collins GS, et al. A systematic review shows no performance benefit of machine learning

over logistic regression for clinical prediction models. J Clin Epidemiol (2019).

30. Ennis M, Hinton G, Naylor D, et al. A comparison of statistical learning methods on the GUSTO database. Stat Med. (1998).

31. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. New England Journal of Medicine. 2019.

32. Kononenko I. Machine learning for medical diagnosis: History, state of the art and perspective. Artif Intell Med. (2001)

33. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: A large-scale benchmark experiment. BMC Bioinformatics 11 (2018): 23-29.