Human Reproduction, Vol.28, No.5 pp. 1280–1288, 2013
Advanced Access publication on March 5, 2013 doi:10.1093/humrep/det017

human reproduction

ORIGINAL ARTICLE *Infertility*

External validation of the endometriosis fertility index (EFI) staging system for predicting non-ART pregnancy after endometriosis surgery

C. Tomassetti[1,*], B. Geysenbergh[1], C. Meuleman[1], D. Timmerman[2], S. Fieuws[3,4], and T. D'Hooghe[1]

[1]KU Leuven, Leuven University Fertility Center, University Hospitals Leuven, 49 Herestraat, Leuven 3000, Belgium [2]KU Leuven, Department Obstetrics and Gynaecology, University Hospitals Leuven, Leuven 3000, Belgium [3]KU Leuven, Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Leuven 3000, Belgium [4]Universiteit Hasselt, Diepenbeek, Belgium

*Correspondence address. E-mail: carla.tomassetti@uzleuven.be

Submitted on July 18, 2012; resubmitted on December 11, 2012; accepted on January 11, 2013

STUDY QUESTION: Can the ability of the endometriosis fertility index (EFI) to predict non-assisted reproductive technology (ART) pregnancy after endometriosis surgery be confirmed by an external validation study?

SUMMARY ANSWER: The significant relationship between the EFI score and the time to non-ART pregnancy observed in our study represents an external validation of this scoring system.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS: The EFI was previously developed and tested prospectively in a single center, but up to now no external validation has been published. Our data provide validation of the EFI in an external fertility unit on a robust scientific basis, to identify couples with a good prognosis for spontaneous conception who can therefore defer ART treatment, regardless of their revised American Fertility Society (rAFS) endometriosis staging.

DESIGN: Retrospective cohort study where the EFI was calculated based on history and detailed surgical findings, and related to pregnancy outcome in 233 women attempting non-ART conception immediately after surgery; all data used for EFI calculation and analysis of reproductive outcome had been collected prospectively as part of another study.

PARTICIPANTS AND SETTING: The EFI score was calculated (score 0–10) for 233 women with all rAFS endometriosis stages (minimal–mild, $n = 75$; moderate–severe, $n = 158$) after endometriosis surgery (1 September 2006–30 September 2010) in a university hospital-based reproductive medicine unit with combined expertise in reproductive surgery and medically assisted reproduction. All participants attempted non-ART conception immediately after surgery by natural intercourse, ovulation induction with timed intercourse or intrauterine insemination (with or without ovulation induction or controlled ovarian stimulation).

DATA ANALYSIS METHOD: All analyses were performed for three different definitions of pregnancy [overall (any HCG >25 IU/l), clinical and ongoing >20 weeks]. Six groups were distinguished (EFI scores 1–3, 4, 5, 6, 7+8, 9+10), and Kaplan–Meier (K–M) estimates for cumulative pregnancy rate were calculated. Subjects were censored when they were lost to follow-up, had subsequent surgery for endometriosis, started ovarian suppression or underwent ART. As K–M estimates might overestimate the actual event rate, cumulative incidence estimates treating ART as competing event were also calculated. Cox regression analysis was used to assess the performance of EFI and constituting variables. Performance of the score (prediction, discrimination) was quantified with the following methods: mean squared error of prediction (Brier score), areas under the receiver-operating curve and global concordance index C^{τ}.

MAIN RESULTS AND THE ROLE OF CHANCE: There was a highly significant relationship between the EFI and the time to non-ART pregnancy (cumulative overall pregnancy rate, $P = 0.0004$), with the K–M estimate of cumulative overall pregnancy rate at 12 months after surgery equal to 45.5% [95% confidence interval (CI) 39.47–49.87]—ranging from 16.67% (95% CI 5.01–47.65) for EFI scores 0–3, to 62.55% (95% CI 55.18–69.94) for EFI scores 9–10. For each increase of 1 point in the EFI score, the relative risk of becoming pregnant increased by 31% (95% CI 16–47%; i.e. hazard ratio 1.31). The 'least function score'—which assesses the tubal/ovarian function at

conclusion of surgery—was found to be the most important contributor to the total EFI score among all the other variables (age, duration of infertility, prior pregnancy, AFS endometriosis lesion and total score).

BIAS, CONFOUNDING AND OTHER REASONS FOR CAUTION: The EFI score had a moderate performance in the prediction of the pregnancy rate. Indeed, the decrease in prediction error was rather small, as shown by the decrease in Brier score from 0.213 to 0.198, and low estimates for R^2 (13%) and C^{τ} (0.629).

GENERALIZABILITY TO OTHER POPULATIONS: As the EFI was validated externally in our own European population after initial testing by Adamson and Pasta (Endometriosis fertility index: the new, validated endometriosis staging system. *Fertil Steril* 2010;**94**:1609-1615) in an American population, it appears that the EFI can be used clinically to counsel infertile endometriosis patients receiving reproductive surgery in specialized centers about their post-operative conception options.

STUDY FUNDING/COMPETING INTEREST(S): This research was supported by funds obtained via the Clinical Research Fund of the University Hospitals Leuven, Belgium, via the Ferring Chair in Reproductive Medicine and Surgery, and the Serono Chair in Reproductive Medicine granted to the Leuven University Fertility Center. The authors have no conflicts of interest to declare.

Key words: endometriosis / laparoscopy / fertility / cumulative pregnancy rate / endometriosis fertility index

Introduction

In women suffering from endometriosis and infertility, the decision as to when and how to perform surgical excision and/or fertility treatment [including intrauterine insemination (IUI) and assisted reproductive technology (ART)] is mainly based on clinical guidelines and expert opinions, like the endometriosis guidelines of the European Society for Human Reproduction and Embryology (ESHRE, Kennedy *et al*., 2005), the recommendations of the Practice Committee of the ASRM (The Practice Committee of the American Society for Reproductive Medicine, 2006) and proposed clinical algorithm (De Ziegler *et al*., 2010; Vercellini *et al*., 2009), rather than on statistical prediction models. These guidelines and opinions are based on literature data such as the Cochrane review (Jacobson *et al*., 2010), which concludes that surgical treatment can be effective for subfertility associated with minimal–mild endometriosis, based on two randomized controlled trials with conflicting results (Marcoux *et al*., 1997; Parazzini, 1999). Preliminary data also suggest that complete surgical removal of minimal to mild endometriosis before the start of ART can improve reproductive outcome following ART treatment (Opøien *et al*., 2011). Other non-randomized studies suggest that complete removal of deeply infiltrating endometriosis potentially improves fertility as well (Chapron *et al*., 1999; Redwine and Wright, 2001; Abbott *et al*., 2003; Daraï *et al*., 2005; Meuleman *et al*., 2009; Meuleman *et al*., 2011b, 2013). However, so far data from randomized controlled trials or meta-analyses to answer the question whether surgical treatment of moderate to severe endometriosis can indeed enhance pregnancy rates compared with expectant management are lacking, as not all studies report fertility outcome or supply sufficiently detailed information (Meuleman *et al*., 2011a).

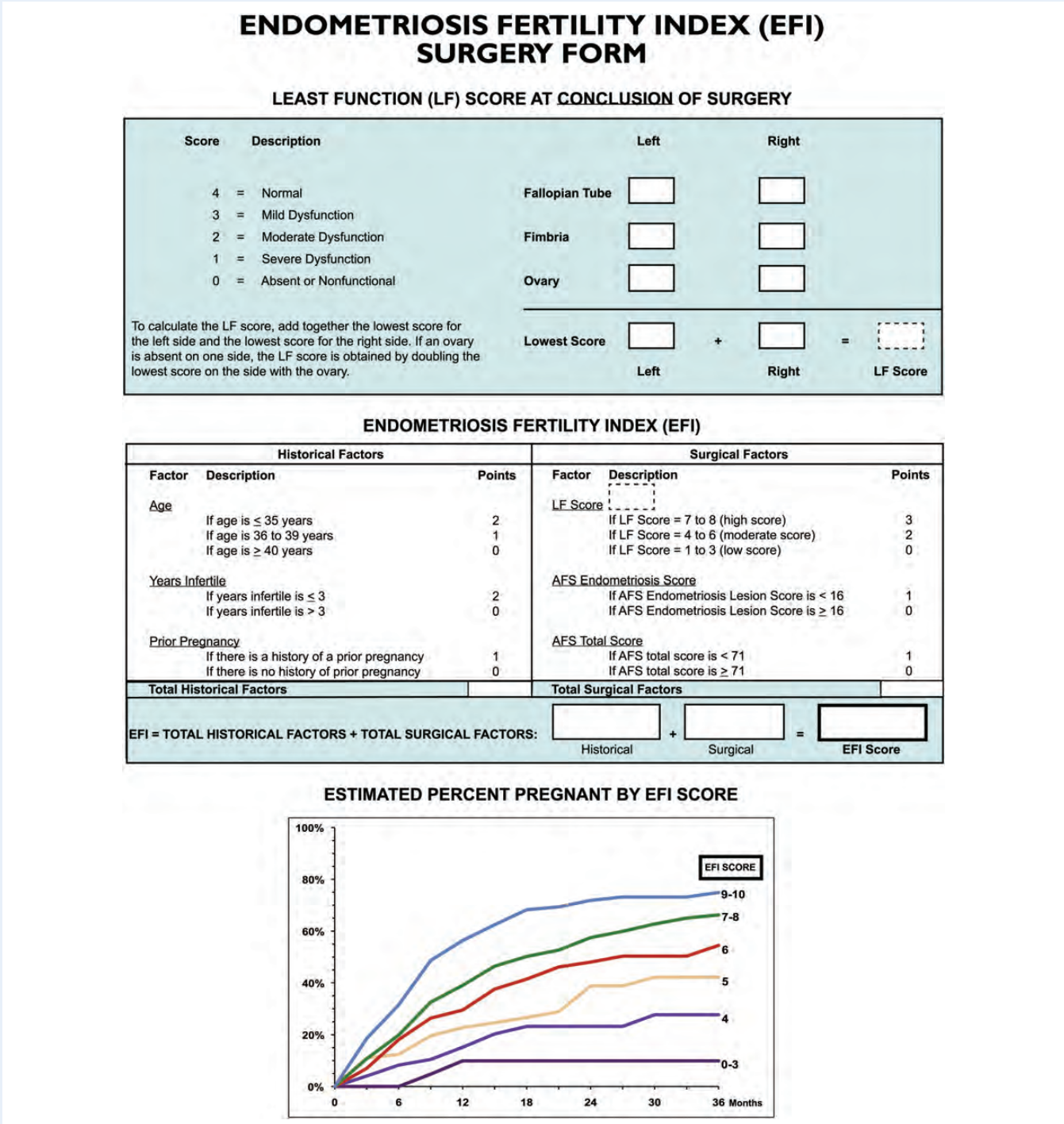
To date, the most frequently used staging system for endometriosis is the revised American Fertility Society (rAFS) score (ASRM, 1997). Unfortunately, this classification system has some serious limitations, including not effectively predicting clinical outcomes of treatment, especially pregnancy rates in infertile patients (Palmisano *et al*., 1993; Vercellini *et al*. 2006; Adamson 2011). For this reason, Adamson and Pasta (2010) developed the endometriosis fertility index (EFI). As shown in Fig. 1, the EFI is a scoring system which includes assessment of historical factors at the time of surgery (age, duration of infertility and pregnancy history), of adnexal function at conclusion of surgery (functional score of Fallopian tubes, fimbriae and ovaries bilaterally), and of the extensiveness of endometriosis (rAFS endometriosis lesion score and total rAFS score). The EFI was established by identifying the most predictive variables for pregnancy from prospectively collected clinical data from 579 patients living in the USA, and subsequently validated by the same authors by testing the correlation of predicted and actual outcomes of an additional 222 patients from the USA who were followed up prospectively (Adamson and Pasta, 2010). The EFI is intended as a clinical tool to counsel patients on the approach towards fertility after surgery. However, up to now, no articles have been published in the English language to validate the EFI in other (external) populations than the original study population mentioned above (Adamson and Pasta, 2010). Such validation is important to confirm the clinical value of the EFI. Therefore, the aim of this study was to provide an external validation of the EFI with robust statistical analysis in a population of infertile endometriosis patients in Belgium.

Materials and Methods

Study population: patient selection

The protocol of this validation study (ClinicalTrials.gov ID: NCT00463398) was approved by the Ethical Committee of the University Hospital Leuven, Belgium as part of a larger prospective follow-up study in patients receiving endometriosis surgery and giving informed consent (Meuleman *et al*., 2013).

From 1 September 2006 onwards, data on history, surgery and detailed post-operative short and long-term outcome (including fertility outcome) of all endometriosis patients treated at the Leuven University Fertility Centre (LUFc) were collected prospectively. All patients were operated at a single site, namely the LUFc at Campus Gasthuisberg of the University Hospitals of Leuven, Belgium. Patients from 'satellite centers' were not included in this analysis. The LUFc is a tertiary referral center for endometriosis, reproductive endocrinology and reproductive surgery (Debrock *et al*., 2010; Meuleman *et al*., 2013). The surgical approach for treatment of endometriosis with CO_2 laser laparoscopy at LUFc has already been described elsewhere (Meuleman *et al*., 2009). Post-operative fertility management of these patients was depending on age and other reproductive factors (as described in Meuleman *et al*., 2013). In women younger than 36–38 years with a regular ovulatory cycle, at least unilaterally normal



ENDOMETRIOSIS FERTILITY INDEX (EFI)
SURGERY FORM

LEAST FUNCTION (LF) SCORE AT CONCLUSION OF SURGERY

Score		Description		Left	Right		
4	=	Normal	Fallopian Tube				
3	=	Mild Dysfunction					
2	=	Moderate Dysfunction	Fimbria				
1	=	Severe Dysfunction					
0	=	Absent or Nonfunctional	Ovary				
			Lowest Score		+		=
				Left		Right	LF Score

To calculate the LF score, add together the lowest score for the left side and the lowest score for the right side. If an ovary is absent on one side, the LF score is obtained by doubling the lowest score on the side with the ovary.

ENDOMETRIOSIS FERTILITY INDEX (EFI)

Historical Factors			Surgical Factors		
Factor	Description	Points	Factor	Description	Points
Age			LF Score		
	If age is ≤ 35 years	2		If LF Score = 7 to 8 (high score)	3
	If age is 36 to 39 years	1		If LF Score = 4 to 6 (moderate score)	2
	If age is ≥ 40 years	0		If LF Score = 1 to 3 (low score)	0
Years Infertile			AFS Endometriosis Score		
	If years infertile is ≤ 3	2		If AFS Endometriosis Lesion Score is < 16	1
	If years infertile is > 3	0		If AFS Endometriosis Lesion Score is ≥ 16	0
Prior Pregnancy			AFS Total Score		
	If there is a history of a prior pregnancy	1		If AFS total score is < 71	1
	If there is no history of prior pregnancy	0		If AFS total score is ≥ 71	0
Total Historical Factors			Total Surgical Factors		

EFI = TOTAL HISTORICAL FACTORS + TOTAL SURGICAL FACTORS: Historical + Surgical = EFI Score

ESTIMATED PERCENT PREGNANT BY EFI SCORE

Figure 1 EFI surgery form and a simplified figure for patient education showing the estimated pregnancy chance per EFI score. Reprinted from *Fertility and Sterility*, 94/5, Adamson GD and Pasta DJ, Endometriosis fertility index: the new, validated endometriosis staging system, 1609–1615, Copyright 2010, with permission from Elsevier.

tubal function and a male partner with a normal sperm analysis, spontaneous conception was expected within 6 months to 1 year. In case of failed spontaneous conception or at least unilaterally normal tubal function with ovulatory dysfunction or mild male factor infertility, controlled ovarian stimulation combined with IUI was performed during 3–4 (maximally 6) cycles. If this treatment was not successful, ART was proposed. In all cases of compromised ovum pick-up and transport capacity and/or major sperm problems and/or advanced female age (>36–38 years), ART was proposed immediately after the patient had recovered from the intervention. None of the fertility management decisions were based on the patients' EFI score as at that moment EFI-scores were not available (EFI was not yet published).

From the previously mentioned prospectively maintained database were selected all infertile patients who were operated for any stage of

endometriosis between 1 September 2006 and 30 September 2010. This time period was chosen to have sufficient length of follow-up to be able to report on fertility. All selected patients wanted to become pregnant immediately after surgery ($n = 326$). Patients proceeding with ART (including IVF and ICSI but excluding IUI, as in WHO/ICMART-definitions by Zegers-Hochschild *et al*., 2009) immediately post-operatively without attempting non-ART pregnancy between surgery and ART were excluded from analysis, thus leaving 233 patients eligible for validation analysis. In December 2011, the EFI was easily calculated based on the highly detailed information found in the prospectively collected database, which contained all required data. The primary outcome measure was the cumulative non-ART pregnancy rate after endometriosis surgery. Pregnancy was defined in three different ways based on WHO–ICMART definitions (Zegers-Hochschild *et al*., 2009): overall pregnancy (any HCG > 25 IU/l), clinical pregnancy (pregnancy diagnosed by ultrasound visualization of one or more gestational sacs or definitive clinical signs of pregnancy, including ectopic pregnancy) and ongoing pregnancy reaching 20 weeks or more [based on (calculated) last menstrual period]. Non-ART pregnancies were defined as conceptions occurring spontaneously, after ovulation induction with timed intercourse, and after IUI (with or without ovulation induction or controlled ovarian stimulation). In patients who needed IUI with donor sperm (because of a partner with azoospermia, $n = 3$) post-operative follow-up time in months was defined as being equal to the number of post-operative IUI-donor-cycles.

Statistical methodology

The EFI was calculated as demonstrated in Fig. 1, which shows how the scoring system works and the form that can be filled in at the end of surgery. The EFI scores range from 0 to 10, with 0 representing the poorest prognosis and 10 the best prognosis for future non-ART pregnancy.

Analyses were performed separately for the three different definitions of pregnancy as described above. To enhance comparability with the original EFI paper (Adamson and Pasta, 2010) and allow external validation, six groups were distinguished based on the EFI score, and life table analysis [Kaplan–Meier (K–M) estimates] was used to construct a curve for the cumulative pregnancy rate. Subjects were censored when they were lost to follow-up, had subsequent surgery for recurrent endometriosis, started ovarian suppression medication or underwent ART. However, these K–M estimates, using the censoring rules as described above, overestimate the non-ART pregnancy rate (Daya, 2005) which can be expected in real practice, because they assume that patients undergoing ART can still become pregnant through a non-ART approach. Therefore, cumulative incidence estimates (Pintilie, 2006) were also calculated next to K–M estimates as they consider that patients undergoing ART have no chance of becoming pregnant with a non-ART approach.

As the usefulness of a scoring system in clinical practice should not be based only on statistical significance, the performance (predictive accuracy and discriminative ability) of the EFI score was also quantified. For continuous and/or binary response models such quantification is common practice, with indices as mean squared error (MSE), proportion of variation explained by the model (R^2) and the area under the receiving-operating curve (AUC), the latter resulting from values for sensitivity and specificity and also known as the index of concordance (*C*-index). For survival data, extensions of these indices have been proposed more recently (Heagerty and Zheng, 2005).

The MSE of prediction (the so-called Brier score) is estimated by a weighted average of time-dependent residuals, with the residual defined as the difference of the time-dependent survival status (pregnant/not pregnant) and the predicted probability from a Cox regression model (Gerds and Schumacher, 2006). This score function is averaged over a follow-up period yielding an overall Brier score. If the prediction of the status at each point during follow-up were perfect for each woman, the Brier score would equal zero. To further ease interpretation of the Brier score, one should consider in our population with prevalence of non-ART pregnancy around 50%, a random prediction would generate a Brier score of 0.25.

An R^2 based on a likelihood of the Cox regression (Royston, 2006) was also calculated.

Further, AUC were obtained at different time points (Heagerty and Zheng, 2005). The time-dependent AUC has a similar interpretation as the AUC for binary responses. It is the probability that the predictions (e.g. a linear function of the EFI score) for a random pair of subjects are concordant with their outcomes (i.e. their status at the specific time point). The AUC ranges between 0.5 (random prediction) and 1 (perfect discrimination). The time-dependent AUC values were combined into a global concordance index C^{τ}. The quantifications were evaluated for the EFI score (whether or not grouped into the same subsets that were proposed by Adamson and Pasta, 2010) and the constituting variables. The χ^2 values and the degrees of freedom (df) of the various models were reported. Restricted cubic splines with five knots (Harrell, 2001) were used to verify if a linear function was sufficient to describe the relationship between the EFI score and the (log) relative risk in the Cox model.

Analyses were performed using SAS software, version 9.2 of the SAS System for Windows (SAS Institute Inc., Cary, NC, USA). The function *risksetAUC* in the R package was used to obtain time-dependent AUC curves and the global concordance indices C^{τ}. The prediction error (Brier score) was calculated with the R function *PEC*.

Results

Characteristics of the 233 patients are shown in Table I, where variables that were used to calculate the EFI are marked with an asterisk. Figure 2A shows the distribution of the EFI score amongst all included patients; the median EFI score was 8. The median EFI score in the group of excluded patients who went directly to ART was 6. As the majority of reported non-ART pregnancies (98/108) reached duration of 20 weeks or more, and only a single biochemical pregnancy was observed, results of the three analyses largely overlap. For this reason, only results for the overall pregnancy rate (any HCG > 25 IU/l) are reported in this article. This is assumed to be in congruence with the original article of Adamson and Pasta (2010), where 'pregnancy' was not defined in more detail. The results for 'clinical pregnancy' and '>20 weeks pregnancy' can be found in the Supplementary data Section. It is interesting to note that 64% (69/108) pregnancies occurred spontaneously (without ovulation induction or controlled ovarian stimulation), whereas 36% (39/108) occurred after treatment with IUI.

The K–M estimates slightly overestimated the actual non-ART pregnancy rate, when compared with the cumulative incidence (CI) estimates (Fig. 2B). For example, for the whole population the 1-year non-ART pregnancy rate equaled 44.5% [95% confidence interval (CI) 39.5–49.9%] and 41.4% (95% CI 34.9–48.0%) for K–M and CI estimates, respectively. After 6 months, these estimates equaled 32.2% (95% CI 27.4–37.6%) and 31.5% (95% CI 25.4–37.6%), respectively.

A highly significant relation was observed between the EFI and the time to non-ART pregnancy ($P = 0.0004$, Fig. 2C). Due to smaller sample sizes for the lower EFI score (Fig. 2A), the curves show less discrimination in these classes. Although the relationship between

Table I Description of the study population: patient characteristics (total population = 233); variables marked with an asterisk are used to calculate the EFI.

	N/total (%) unless shown otherwise
Age at surgery (years) Mean ± SD	31.3 ± 3.9
≤35 year*	202/233 (87)
36–39 year*	29/233 (125)
≥40 year*	2/233 (16)
At least one previous therapeutic surgery for endometriosis[a]	115/233 (491)
At least one previous failed IUI cycle	46/233 (20)
At least one previous failed ART cycle	22/233 (9)
Duration of infertility (months) Median (min–max)	24 (1–120)
>3 years*	56/233 (24)
≤3years*	177/233 (76)
Prior pregnancy	
No*	174/233 (75)
Yes*	59/233 (25)
Least function score[b],* Mean ± SD	5.7 ± 1.8
High score (7–8)*	85/233 (36)
Moderate score (4–6)*	121 (52)
Low score (1–3)*	27/233 (12)
AFS endometriosis lesion score[c] Median (min–max)	10 (1–52)
<16*	155/233 (67)
≥16*	78/233 (337)
AFS total score[d] Median (min–max)	36 (1–126)
<16	75/233 (327)
≥16	158/233 (68)
<71*	190/233 (82)
≥71*	43/233 (18.45)
Partner with normospermia[e]	106/197 (54)
TMC[f] of partner's sperm Median (min–max; $n = 189$)	20 (0–117)

IUI, Intrauterine insemination; ART, Assisted reproductive technology.
[a]Excluding mere diagnostic surgery.
[b]See Table I for calculation.
[c]Revised American Fertility Society classification: total of the endometriosis lesion score.
[d]Revised American Fertility Society classification: total score (including endometriosis lesion, cul-de-sac and adhesion score).
[e]Based on WHO-classification and strict morphology criteria (WHO, 1999).
[f]Total motile sperm count after sperm preparation (capacitation).

EFI and time to non-ART pregnancy was highly significant, the general performance (predictive accuracy, decrease in prediction error) of the EFI to predict non-ART pregnancy was only moderate, as quantified by the following analyses. First of all, low-to-moderate estimates for the R^2 and C^τ are obtained, i.e. 0.13 and 0.629. The R^2 of 0.13 can be interpreted as 13% of the difference between patients is being explained by the EFI. Furthermore, the expected Brier score also only decreases from 0.213 in the whole population (based on the K–M estimates, without taking into account the EFI scores) to 0.198 with taking into account the EFI score (Table II). Although at the extremes of the EFI range the effect was less strong, linearity in the effect of EFI was not rejected ($P = 0.76$). The chance for a non-ART pregnancy increases relatively for each point with 31% (95% CI 16–47%; i.e. hazard ratio of 1.31).

Table II shows that the surgical factors were more important than the historical factors in predicting overall pregnancy rate according to EFI. The end-of-surgery least function (LF) score was the most important contributor to the EFI, as revealed by the comparison of the χ^2 values of the various Cox models. This is shown in Table II, where the χ^2 value of the model with the EFI LF (three levels) only drops to 18.4 (df = 2), compared with an χ^2 of 24.0 (df = 5) of the model using the classification of the EFI total score; the R^2 equals 10% and the C^τ evaluated over the first 24 months 0.60. However, there is no evidence that the information not captured in the EFI LF score is redundant. This was confirmed by using a (bivariable) Cox regression model with as predictors the EFI LF score and a 'rest' score (being the difference between the EFI total score and the EFI LF score). The latter score still had a significant contribution ($\chi^2 = 5.8$, df = 1, $P = 0.016$). When considering the AFS total score as a contributor to the EFI (score of more than 71), Table II shows that this also is an important part of the EFI, but the EFI itself performs better as illustrated by the low R^2; a similar conclusion can be drawn for the contribution of the endometriosis lesion score.

Discussion

In our study of 233 women attempting non-ART conception immediately after therapeutic endometriosis surgery study, we were able to externally validate the EFI, which had originally been proposed as a clinical tool to predict pregnancy rates after surgery for endometriosis (Adamson and Pasta, 2010). This is shown in Fig. 2C, illustrating the significant relationship between the EFI and the time to non-ART pregnancy. To our knowledge, this is the first external validation and therefore provides important information. It needs to be stressed that surgery was primarily performed laparoscopically aiming at complete excision of all endometriotic lesions using CO_2-laser, with maximum respect for the preservation and/or reconstruction of reproductive anatomy. Adding to its importance is the fact that our study might be more rigorous than the original study (Adamson and Pasta, 2010) because the performance of the EFI was evaluated and reported on in the present study. Clearly, the differentiation in pregnancy rates between the EFI classes follows the expected pattern more closely amongst the higher EFI scores, but this might be due to the smaller sample sizes for patient groups with lower EFI scores (Fig. 2A) in our population, as in the population studied in the original publication (Adamson and Pasta, 2010). The latter is a reflection of current routine clinical practice, where women with intuitively assumed bad prognosis for non-ART pregnancy (e.g. severe bilateral tubal damage observed at surgery, advanced female age; Meuleman *et al.*, 2013) are immediately referred for ART without attempting natural conception or IUI. This obviously is a form of 'active censorship' that cannot be ruled out as bias, as up to now the decision on how to counsel a post-operative endometriosis patients did not depend on the EFI score. The median EFI score appeared to be higher in our study group (median EFI of 8) than in the excluded group of women who

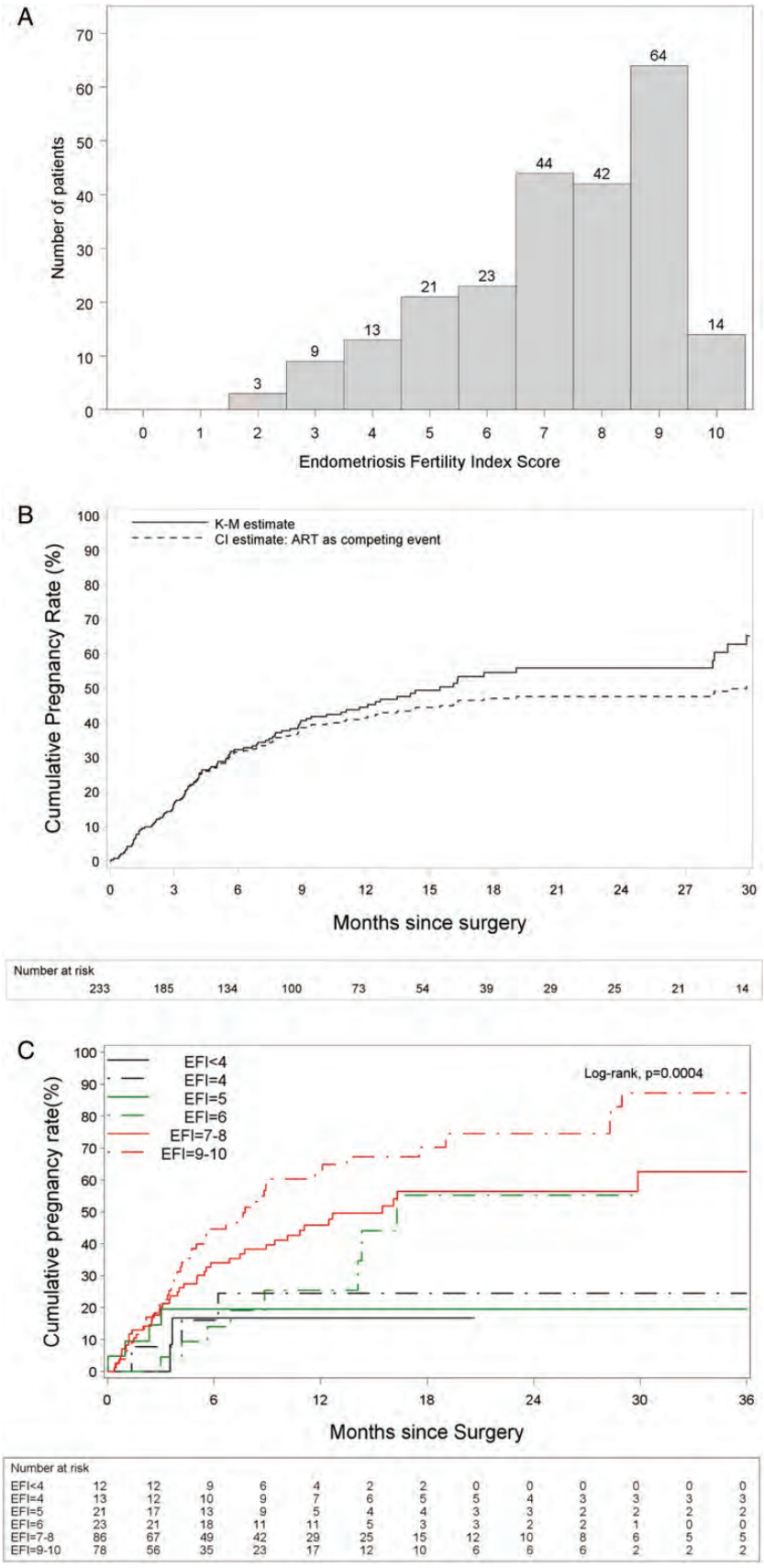


Figure 2 (**A**) Distribution of patients according to their EFI score. (**B**) Comparison of K–M estimates versus CI estimates for the total group. (**C**) K–M estimates for the cumulative pregnancy rate as a function of the EFI.

Table II Overview of the performance of various factors predicting overall pregnancy rate according to EFI.

Factor in cox model	χ^2	df	P	MSE_0	MSE_{model}	R^2	C^τ
EFI continuous (linear)[a]	22.49	1	<0.001	0.213	0.198	0.12	0.627
EFI Adamson[b]	24.01	5	<0.001	0.213	0.198	0.13	0.626
All historical factors	12.76	5	0.026	0.213	0.203	0.07	0.576
Age	7.43	2	0.024	0.213	0.208	0.04	0.538
Years infertile	2.58	1	0.108	0.213	0.211	0.01	0.532
Prior pregnancy	4.97	1	0.026	0.213	0.209	0.03	0.546
All surgical factors	20.89	5	<0.001	0.213	0.202	0.11	0.600
Least function score(groups)[b]	18.35	2	<0.001	0.213	0.202	0.10	0.600
Least function score (linear)[a]	20.86	1	<0.001	0.213	0.199	0.11	0.624
AFS Endometriosis score	2.12	1	0.146	0.213	0.211	0.01	0.534
AFS Total score	8.18	1	0.004	0.213	0.209	0.05	0.553

χ^2= chi-square value from a Cox regression model. P= P-value likelihood-ratio test. df, degrees of freedom. MSE_0= Mean-squared error (Brier score) from a prediction using no EFI information at all. The prediction is the same for all women and obtained from the K–M curve for the total sample. MSE_{model}= Brier score obtained with a prediction from the Cox regression with a specific predictor. R^2: R^2-like quantification of predictive accuracy (Royston, 2006), i.e. reflecting a proportion of explained variance. C^τ: global concordance index (AUC) defined over a follow-up period of 2 years (Heagerty and Zheng, 2005). A value equal to 0.5 refers to random prediction, whereas C^τ= 1 implies perfect discrimination between pregnant and non-pregnant.

[a]Each score considered individually from 0 to 10.

[b]With grouping of scores 0–3, 4, 5, 6, 7–8 and 9–10.

went directly post-operatively to ART without attempting non-ART conception (median EFI of 6), suggesting that our clinical post-operative judgment was associated with selection of 'better prognosis patients' for non-ART conception, even though at the time of decision-making their actual EFI score was not known (confer retrospectively calculated).

Next to this, our study included more patients with a high EFI versus a moderate EFI score when compared with the original publication (Adamson and Pasta, 2010), i.e. more patients who had a higher potential to be pregnant. However, we do not believe that this difference would have led to an overestimation of the EFI score validity, as a linear relationship was shown between the EFI score and the pregnancy rate. If the fact that more 'good prognosis' patients are included would skew the data favorably, a non-linearity would be observed, with a proportionally higher raise in pregnancy rate per higher EFI point, which was not observed in our study.

Our data further confirm that the end-of-surgery LF score is the most important contributor to the EFI score, without evidence that the information not captured in the EFI LF score is redundant. Again, the importance of the adnexal function is already intuitively used in guidelines such as the ESHRE-guidelines (Kennedy *et al.*, 2005) as well as in our own clinical practice (Meuleman *et al.*, 2011b), where patients with problematic tubal function or already diminished ovarian reserve are offered ART immediately after surgery. A possible criticism of the EFI score is that it includes some dependent parameters such as various values of the ASRM score. Adamson and Pasta (2010) state in their article that, after identifying factors in their database being most predictive for pregnancy and addressing the importance of a group of variables, subsequent analyses combined predictive variables and in that way established the EFI scoring system; they do not comment on inter-dependence of all variables (Adamson and Pasta, 2010). However, as the aim of this study was not to improve or alter the EFI score, but merely to validate it externally as it was published, it is clear that we cannot account for this.

Although the relationship between EFI and non-ART pregnancy rate is highly significant, the general performance of the EFI (predictive accuracy, discriminative ability) is moderate. Unfortunately, no information on the performance of the EFI score is given in the original publication of Adamson and Pasta (2010), as they only report a 'good correlation of predicted and actual outcomes for all stages of endometriosis', therefore direct comparison with our data is not possible. Obviously, the moderate predictive and discriminative performance of the EFI may attenuate the clinical importance of the EFI as a scoring system to educate patients. This is, however, not unusual, as the general performance of many scoring systems has frequently been found to be low or moderate by survival analyses (Steyerberg *et al.*, 2010). This does not imply that the scoring system is without value, because even a poorly discriminating model may be clinically useful if the clinical decision is close to a 'toss-up', because it implies that the threshold is in the middle of the distribution of predicted risks, as is the case for models in fertility medicine for example (Steyerberg *et al.*, 2010).

A possible limitation of our study is its retrospective study nature. It needs to be stressed, however, that all data were acquired in a meticulously maintained prospective database as part of a larger prospective study (Meuleman *et al.*, 2013) including nearly all patients operated for endometriosis in our center in the time period of 2006–2010. The EFI score was calculated exclusively based on these prospectively collected data. So far, other staging and scoring systems except the EFI have previously failed to show any correlation with post-operative fertility (Palmisano *et al.* 1993, Adamson 2011). Therefore, in our opinion, the EFI is the best (and only) objective scoring system available to guide the counseling of patients after surgery about their fertility prognosis and eventual need for fertility treatment. The fact

that the validation was successful in our dataset—albeit in a different continent—in our opinion greatly increases the clinical value of the EFI. Clinical use of the EFI in real practice can provide a basis for deferring treatment with ART in the better prognostic groups regardless of the rAFS endometriosis staging. As the EFI also takes into account maternal age, and the mean age of our patients was relatively young (31.3 years), it is possible that patients with better prognosis will be younger, which in turn is beneficial to allow time to wait for a natural conception. If in turn a patient with a good score has a higher age, the time of spontaneous evolution might be reduced to prevent loss of ovarian reserve. In our study, female age did not significantly impact post-operative fertility management decisions, as it was comparable in the study group (mean 31.3 SD $\pm$ 3.9 years) and in the group of patients excluded from our study because of direct post-operative assignment to ART without any non-ART attempts ($n = 93$, mean 32.2 years, SD $\pm$ 3.8 years). Based on our data, it is therefore impossible to make any firm recommendations until which age which time for spontaneous conception should be allowed in patients with high EFI scores. In our opinion, future research should concentrate on assessing and refining the performance of the EFI as a triage instrument for decision making in post-operative fertility management in endometriosis patients. More research is also needed to confirm the interesting observation that complete surgical removal of minimal-to-mild endometriosis before the start of ART can improve reproductive outcome following ART treatment (Opøien *et al.*, 2011).

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Authors' roles

C.T., B.G., C.M. and T.D. designed the study; C.T. and B.G. acquired and processed data; C.T., S.F. and T.D. analyzed data; C.T. and B.G. drafted the manuscript; C.T., C.M., D.T. and T.D. were involved in the critical discussion; C.T. finalized the manuscript.

Funding

This research was supported by funds the Clinical Research Fund of the University Hospitals Leuven, Belgium. The Leuven University Fertility Centre also receives unrestricted research grants via the Ferring Chair in Reproductive Medicine and Surgery, and the Serono Chair in Reproductive Medicine.

Conflict of interest

None declared.

References

Abbott JA, Hawe J, Clayton RD, Garry R. The effect and effectiveness of laparoscopic excision of endometriosis: a prospective study with 2–5 year follow-up. *Hum Reprod* 2003;**18**:1922–1927.

Adamson GD. Endometriosis classification: an update. *Curr Opin Obstet Gynecol* 2011;**23**:213–220.

Adamson GD, Pasta DJ. Endometriosis fertility index: the new, validated endometriosis staging system. *Fertil Steril* 2010;**94**:1609–1615.

American Society for Reproductive Medicine. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertil Steril* 1997;**67**:817–821.

Chapron C, Fritel X, Dubuisson JB. Fertility after laparoscopic management of deep endometriosis infiltrating the uterosacral ligaments. *Hum Reprod* 1999;**14**:329–332.

Daraï E, Marpeau O, Thomassin I, Dubernard G, Barranger E, Bazot M. Fertility after laparoscopic colorectal resection for endometriosis: preliminary results. *Fertil Steril* 2005;**84**:945–950.

Daya S. Life table (survival) analysis to generate cumulative pregnancy rates in assisted reproduction: are we overestimating our success rates? *Hum Reprod* 2005;**20**:1135–1143.

Debrock S, Melotte C, Spiessens C, Peeraer K, Vanneste E, Meeuwis L, Meuleman C, Frijns JP, Vermeesch JR, D'Hooghe TM. Preimplantation genetic screening for aneuploidy of embryos after in vitro fertilization in women aged at least 35 years: a prospective randomized trial. *Fertil Steril* 2010;**93**:364–73.

De Ziegler D, Borghese B, Chapron C. Endometriosis and infertility: pathophysiology and management. *Lancet* 2010;**376**:730–738.

Gerds TA, Schumacher M. Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biom J* 2006;**48**:1029–1040.

Harrell FE. *Regression Modeling Strategies*. 2001, New York: Springer.

Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005;**61**:92–105.

Jacobson TZ, Duffy JM, Barlow D, Farquhar C, Koninckx PR, Olive D. Laparoscopic surgery for subfertility associated with endometriosis. *Cochrane Database Syst Rev* 2010;**20**:CD001398.

Kennedy S, Bergqvist A, Chapron C, D'Hooghe T, Dunselman G, Greb R, Hummelshoj L, Prentice A, Saridogan E. ESHRE Special Interest Group for Endometriosis and Endometrium Guideline Development Group. ESHRE guideline for the diagnosis and treatment of endometriosis. *Hum Reprod* 2005;**20**:2698–2704.

Marcoux S, Maheux R, Bérubé S. Laparoscopic surgery in infertile women with minimal or mild endometriosis. Canadian Collaborative Group on Endometriosis. *N Engl J Med* 1997;**337**:217–222.

Meuleman C, D'Hoore A, Van Cleynenbreugel B, Beks N, D'Hooghe T. Outcome after multidisciplinary CO_2 laser laparoscopic excision of deep infiltrating colorectal endometriosis. *Reprod Biomed Online* 2009;**18**:282–289.

Meuleman C, Tomassetti C, D'Hoore A, Van Cleynenbreugel B, Penninckx F, Vergote I, D'Hooghe T. Surgical treatment of deeply infiltrating endometriosis with colorectal involvement. *Hum Reprod Update* 2011a;**17**:311–326.

Meuleman C, Tomassetti C, D'Hoore A, Buyens A, Van Cleynenbreugel B, Fieuws S, Penninckx F, Vergote I, D'Hooghe T. Clinical outcome after CO_2 laser laparoscopic radical excision of endometriosis with colorectal wall invasion combined with laparoscopic segmental bowel resection and reanastomosis. *Hum Reprod* 2011b;**26**:2336–2343.

Meuleman C, Tomassetti C, Wolthuis A, Van Cleynenbreugel B, Laenen A, Penninckx F, Vergote I, D'Hoore A, D'Hooghe T. Clinical outcome after radical excision of moderate-severe endometriosis with or without bowel resection and reanastomosis: a prospective cohort study. *Ann Surg* 2013 (in press).

Opøien HK, Fedorcsak P, Byholm T, Tanbo T. Complete surgical removal of minimal and mild endometriosis improves outcome of subsequent IVF/ICSI treatment. *Reprod Biomed Online* 2011;**23**:389–395.

Palmisano GP, Adamson GD, Lamb EJ. Can staging systems for endometriosis based on anatomic location and lesion type predict pregnancy rate? *Int J Fertil Menopausal Stud* 1993;**38**:241–249.

Parazzini F. Ablation of lesions or no treatment in minimal-mild endometriosis in infertile women: a randomized trial. Gruppo Italiano per lo Studio dell' Endometriosi. *Hum Reprod* 1999; **14**:1332–1334.

Pintilie M. *Competing Risks: A Practical Perspective*. Chichester: Wiley, 2006.

Redwine DB, Wright JT. Laparoscopic treatment of complete obliteration of the cul-de-sac associated with endometriosis: long-term follow-up of en bloc resection. *Fertil Steril* 2001;**76**:358–365.

Royston P. Explained variation for survival models. *Stata J* 2006;**6**:83–96.

Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;**21**:128–138.

The practice Committee of the American Society for Reproductive Medicine. Endometriosis and infertility. *Fertil Steril* 2006;**86**:S156–160.

Vercellini P, Fedele L, Aimi G, De Giorgi D, Consonni D, Crosignani PG. Reproductive performance, pain recurrence and disease relapse after conservative surgerical treatment for endometriosis: the predictive value of the current classification system. *Hum Reprod* 2006; **21**:2679–2685.

Vercellini P, Somigliana E, Viganò P, Abbiati A, Barbara G, Crosignani PG. Surgery for endometriosis-associated infertility: a pragmatic approach. *Hum Reprod* 2009;**24**:254–269.

WHO World Health Organization. *Laboratory Manual for the Examination of Human Semen and Sperm-cervical Mucus Interaction*, 4th edn. Cambridge: Cambridge University Press, 1999.

Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, Sullivan E, van der Poel S; International Committee for Monitoring Assisted Reproductive Technology; World Health Organization. The International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) Revised Glossary on ART Terminology. *Hum Reprod* 2009;**24**:2683–2687.