



BMJ Open Predicting the chance on live birth per cycle at each step of the IVF journey: external validation and update of the van Loendersloot multivariable prognostic model

Johanna Devroe ^{1,2}, Karen Peeraer,^{1,2} Geert Verbeke,^{3,4} Carl Spiessens,¹ Joris Vriens,² Eline Dancet ^{1,2,5}

To cite: Devroe J, Peeraer K, Verbeke G, *et al.* Predicting the chance on live birth per cycle at each step of the IVF journey: external validation and update of the van Loendersloot multivariable prognostic model. *BMJ Open* 2020;**10**:e037289. doi:10.1136/bmjopen-2020-037289

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-037289>).

Received 27 January 2020
Revised 06 August 2020
Accepted 11 August 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Johanna Devroe;
Johanna.Devroe@uzleuven.be

ABSTRACT

Objective To study the performance of the ‘van Loendersloot’ prognostic model for our clinic’s in vitro fertilisation (IVF) in its original version, the refitted version and in an adapted version replacing previous by current cycle IVF laboratory variables. **Methods** This retrospective cohort study in our academic tertiary fertility clinic analysed 1281 IVF cycles of 591 couples, who completed at least one 2nd–6th IVF cycle with own fresh gametes after a previous IVF cycle with the same partner in our clinic between 2010 and 2018. The outcome of interest was the chance on a live birth after one complete IVF cycle (including all fresh and frozen embryo transfers from the same episode of ovarian stimulation). Model performance was expressed in terms of discrimination (c-statistics) and calibration (calibration model, comparison of prognosis to observed ratios of five disjoint groups formed by the quintiles of the IVF prognoses and a calibration plot).

Results A total of 344 live births were obtained (26.9%). External validation of the original van Loendersloot model showed a poor c-statistic of 0.64 (95% CI: 0.61 to 0.68) and an underestimation of IVF success. The refitted and the adapted models showed c-statistics of respectively 0.68 (95% CI: 0.65 to 0.71) and 0.74 (95% CI: 0.70 to 0.77). Similar c-statistics were found with cross-validation. Both models showed a good calibration model; refitted model: intercept=0.00 (95% CI: -0.23 to 0.23) and slope=1.00 (95% CI: 0.79 to 1.21); adapted model: intercept=0.00 (95% CI: -0.18 to 0.18) and slope=1.00 (95% CI: 0.83 to 1.17). Prognoses and observed success rates of the disjoint groups matched well for the refitted model and even better for the adapted model.

Conclusion External validation of the original van Loendersloot model indicated that model updating was recommended. The good performance of the refitted and adapted models allows informing couples about their IVF prognosis prior to an IVF cycle and at the time of embryo transfer. Whether this has an impact on couple’s expected success rates, distress and IVF discontinuation can now be studied.

INTRODUCTION

Several groups have developed a prognostic model that combines variables with different weights in order to estimate the probability of in vitro fertilisation (IVF) success.^{1–6} These

Strengths and limitations of this study

- TRIPOD criteria were taken into account to ensure a thorough report on the external validation of one of the best quality prognostic in vitro fertilisation models and on model updating and validation through resampling.
- Performance was expressed in terms of discrimination (ie, c-statistics) and in term of calibration, assessed with several methods.
- The used methodology had the limitation of relying on a retrospective cohort.

models differ in their ability to discriminate between couples having and couples not having success and in their likeliness and extent of ‘miscalibration’ (ie, disagreement between predicted and observed success rates).^{2 4 5 7–11}

Patients and professionals have both proposed to inform couples about their IVF prognosis rather than giving average success rates.^{3 12} Qualitative research showed that women considering to cryopreserve their oocytes expect their chance of success to be higher than average due to, for example, their exceptional lifestyle or their exceptionally skilled gynaecologist.¹³ Couples could be informed prior to an IVF cycle, while they are deciding whether to start an (other) IVF-cycle^{3 12} or at the time of an embryo transfer, while they ask feedback about their cycle.

In order to study the impact of informing couples about their IVF prognosis, a model had to be selected and its performance had to be assessed for our IVF clinic. The ‘van Loendersloot model’ was selected for several reasons.³ First, it is among the three existing prognostic models for which the development has been reported with the highest

quality (as reviewed by Ratna *et al*⁶). Second, on development the van Loendersloot model had a reasonable c-statistic of 0.68 and the goodness-of-fit test showed no significant miscalibration.³ Third, another clinic validated the model successfully.⁹ Fourth, the selected model predicts success rates per complete IVF cycle (including fresh and frozen–thawed embryo transfers from the same episode of ovarian stimulation), which is both relevant for deciding whether to start an (other) IVF cycle and for fostering realistic expectations during an IVF cycle. Fifth, the van Loendersloot model has been validated for calculating the prognosis prior to every IVF cycle and not only for calculating around the first IVF cycle.³ Finally, from the second IVF cycle onwards the model includes IVF laboratory variables from the previous cycle (eg, number of embryos) besides clinical variables (eg, endometriosis), which seems to improve the performance (eg, van Loendersloot *et al* 2013³ vs Templeton *et al* 1996¹). Interestingly, IVF laboratory variables from the current rather than the previous cycle could be included for calculating IVF prognoses at the time of an embryo transfer.

This study aimed to evaluate the performance of the van Loendersloot model for our clinic in its original version and in a refitted version including previous cycle IVF laboratory variables and in an adapted version including current cycle IVF laboratory variables.

MATERIAL AND METHODS

The original van Loendersloot model was externally validated based on a retrospective cohort from the Leuven University fertility clinic.^{14 15} Next, the model was updated to develop the refitted and the adapted model by respectively re-estimating predictor weights for our clinic and by replacing predictors.¹⁵ The validity of both novel models was assessed by relying on resampling from the same cohort with the aid of leave-one-out cross-validation.¹⁶ Data were analysed using the Statistical Package for Social Sciences, V.25.0 (Chicago, Illinois, USA). The TRIPOD checklist was used for writing this manuscript on the external validation of the original van Loendersloot model and on the development and validation using resampling of the two novel prognostic models.¹⁷

This study was approved by the Leuven University and University Hospital ethics committee (s62898). Given the retrospective design, no informed consent was requested but all treated couples provided consent for using their medical data for quality management purposes.

The development and internal validation study³ and the external validation study,⁹ respectively, used ongoing pregnancy and live birth as measure of success per complete IVF cycle. This study focuses on live birth rate (LBR) per complete IVF cycle (ie, at least one live birth after the transfer of fresh and frozen–thawed embryos from the same episode of ovarian stimulation) as we had the required data and agree with the two other groups that live birth is a more ideal measure of success.^{3 9} Live

birth was defined as the delivery of at least one viable newborn after 24 weeks of gestation.^{18 19}

A priori calculations indicated that a sample of at least 714 IVF cycles was required to end up with 200 events as the LBR per episode of ovarian stimulation is about 28%.^{9 20} The target of 200 events is considered a general requirement for validating prognostic models.^{21 22} In addition, it allows sticking to the rule of thumb of at least 10 events for each of the 17 predictors of the van Loendersloot model (ie, including predictor interactions¹⁴).

Eligible couples completed at least one 2nd–6th IVF cycle with their own fresh gametes in our clinic between 2010 and 2018. First IVF cycles were not included in our cohort for three reasons. First, we especially selected the van Loendersloot model as it includes IVF laboratory variables and these are not available prior to a first IVF cycle. Second, our clinical practice relies on the IVF laboratory variables of the first IVF cycle for information on couple's fertility. Third, the pretreatment McLernon model seems more appropriate for clinical use prior to the very start of the IVF journey as it informs couples and clinicians on the LBR after a first cycle and after an entire IVF journey. Couples who did not complete at least one previous IVF cycle with the same partner in our clinic until live birth or until having used all their cryopreserved embryos (ie, no within cycle dropout) were not included as the van Loendersloot model includes previous cycle variables. Couples who were positive for the HIV or who relied on preimplantation genetic testing were excluded. As we planned to perform a complete case analysis, couples were only eligible if all clinical variables, used in the van Loendersloot model, could be extracted from our electronic medical records (EMRs). Next, for all included couples, all IVF laboratory variables were extracted for their 2nd–6th completed stimulated IVF cycles in which a day 3 or 5 transfer was performed (ie, no within cycle dropout, no cancelled cycles, no modified natural cycles and no day 2 transfers). Included cycles were followed up until completion by live birth or by the absence of a live birth, although all embryos from that cycle had been transferred.¹⁴ IVF laboratory variables were also extracted for a previous cycle, as the van Loendersloot model includes previous cycle IVF laboratory variables. There were no missing data at the level of IVF laboratory variables, other than in case of a lack of fertilisation.

During the study period, IVF was performed in line with the Leuven University fertility clinic's protocol. In fresh cycles, controlled ovarian hyperstimulation was carried out with gonadotrophins (regimen and dose based on clinical characteristics) as described previously by Debrock and colleagues.²³ Follicular response was monitored by gynaecological ultrasound and estradiol measurements in peripheral blood. Ultrasound-guided oocyte retrieval was carried out 35 hours after human chorion gonadotrophin (hCG)-injection. Oocyte retrieval, sperm preparation and standard IVF procedures (with or without intracytoplasmic sperm injection) were performed as described by Debrock *et al* (2015).²⁴ About 16–20 hours

post insemination, fertilisation was assessed. On day 2 and 3 embryo development was evaluated according to the number of blastomeres, the degree of fragmentation and the symmetry of the blastomeres.²⁵ Embryo transfer was mostly done on day 3 but exceptionally blastocyst transfer was considered. One or two embryos were transferred according to the Belgian law.^{26 27} Embryo transfer was cancelled if no viable embryo was available or if the patient was at risk of ovarian hyperstimulation syndrome. Luteal supplementation was given by intravaginal application of progesterone (600 mg/day, Utrogestan) as from the evening of hCG injection. Supernumerary embryos of sufficient quality (ie, on day 3: ≥ 6 blastomeres, $\leq 25\%$ fragmentation and symmetry in blastomeres; on day 5: blastocyst formation) were cryopreserved.²⁸ Cryopreservation was performed by slow freezing (2010–2014) or vitrification (2014–2018). Straws were thawed until the number of survived embryos was equal to the number of requested embryos for transfer.²⁹ A maximum of two embryos, that survived thawing ($\geq 50\%$ of intact cells) and resumed mitosis, were replaced as determined by Belgian law.³⁰ Thawed embryos were transferred in natural, stimulated or hormonal replacement cycles.³¹ All oocytes and embryos were cultured in a single medium covered with mineral oil (GM501).

This cohort's clinical and IVF laboratory variables and the chosen measure of success: LBR per complete IVF cycle (ie, at least one live birth after the transfer of fresh and frozen–thawed embryos from the same episode of ovarian stimulation) were extracted simultaneously from EMRs (ie, no blinding).¹⁶ Regarding IVF laboratory variables, both previous and current cycle variables were extracted as this study not only aimed to validate the original van Loendersloot model and to refit it but it also aimed to adapt the model. All clinical and IVF laboratory variables of the refitted and the adapted model were clearly defined (table 1). The cut-offs proposed by van Loendersloot were followed. Whether variables were adapted for each included IVF cycle as they were time sensitive (eg, woman's age) or affected by previous cycles (eg, number of failed IVF cycles) was specified. As van Loendersloot and colleagues³ (2013) did not clearly define the following variables, they were defined in line with our clinical practice: male infertility,³² diminished ovarian reserve,³³ endometriosis and mean morphological quality. The clinical variables include: female age (years), duration of infertility (years), previous delivery (yes/no), male infertility (yes/no; WHO 2010 criteria³¹), diminished ovarian reserve (yes/no; Bologna criteria³³), endometriosis (yes/no; yes=the presence of a

Table 1 The clinical and IVF laboratory variables of the original, the refitted and the adapted van Loendersloot model

	Original and refitted van Loendersloot model for calculating IVF prognoses prior to another IVF cycle	Adapted van Loendersloot model for calculating IVF prognoses at the time of embryo transfer
Clinical variables		Female age (years)‡‡ Duration of infertility (years)*‡‡ Previous delivery (yes/no)‡‡ Male infertility (yes/no)† Diminished ovarian reserve (yes/no)‡ Endometriosis (yes/no)§ Basal FSH (IU/mL)¶ Failed IVF cycles (number)*
IVF laboratory variables	Fertilisation in previous cycle (yes/no) Embryos in previous cycle (number)** Mean MS on day 3 in previous cycle ^{1–4} †† Presence of 8-cell embryos on day 3 in previous cycle (yes/no) Presence of morulae on day 3 in previous cycle (yes/no)	Embryos in current cycle (number)** Mean MS on day 3 in current cycle ^{1–4} †† Presence of 8-cell embryos on day 3 in current cycle (yes/no) Presence of morulae on day 3 in current cycle (yes/no)

*Cut-off: if duration of infertility ≥ 5 years=5 years.

†Categorisation: male infertility was diagnosed based on the WHO criteria.³²

‡Categorisation: diminished ovarian reserve was diagnosed based on the Bologna criteria for diagnosis of poor responders.³³

§Categorisation: endometriosis was based on the presence of a laparoscopic diagnosis of stage III or IV endometriosis.

¶Cut-off: if basal \leq FSH 10 IU/L=10 IU/L.

**Cut-off: if number of embryos ≥ 10 = 10 embryos.

††Categorisation: MS of 1–4 (the lower, the better) for day 3 embryos; MS1: 7, 8 or 9 blastomeres, $<10\%$ fragmentation and equally or approximately equally sized blastomeres ($<50\%$ difference); MS2: more than 7 blastomeres, $<25\%$ fragmentation and equally or approximately equally sized blastomeres ($<50\%$ difference); MS3: 6 blastomeres or >6 blastomeres and 10% – 25% fragmentation in combination with unequally sized blastomeres ($>50\%$ difference); MS4: <6 blastomeres or $>25\%$ fragmentation.

‡‡Clinical variables that were adapted in between the included IVF cycles.

FSH, follicle stimulating hormone; IVF, in vitro fertilisation; MS, morphological score.

laparoscopic diagnosis of stage III or IV endometriosis), basal follicle stimulating hormone (FSH) (IU/mL), number of previous failed IVF cycles (number) and fertilisation in previous cycle (yes/no). The IVF laboratory variables include number of embryos (number), the presence of 8-cell embryos on day 3 (yes/no), the presence of morulae on day 3 (yes/no) and the mean morphological score (MS) on day 3. The latter was calculated by grading¹⁻⁴ (the lower, the better) all day 3 embryos and calculating the mean MS per cycle. The requirements per MS were: MS1: 7, 8 or 9 blastomeres, <10% fragmentation and equally or approximately equally sized blastomeres (<50% difference); MS2: more than 7 blastomeres, <25% fragmentation and equally or approximately equally sized blastomeres (<50% difference); MS3: 6 blastomeres or >6 blastomeres and 10%–25% fragmentation in combination with unequally sized blastomeres (>50% difference); and MS4: <6 blastomeres or >25% fragmentation.

Other than cases missing IVF laboratory variables due to no fertilisation, only complete cases were analysed and the proportion of eligible consecutive couples, which could be included, was reported.¹⁴

Three models were evaluated: the original van Loendersloot model, the refitted van Loendersloot model and the adapted van Loendersloot model. The original and refitted van Loendersloot model calculate IVF prognoses per complete IVF cycle prior to an IVF cycle and derive all five included IVF laboratory variables from the previous cycle. Whereas the original model relies on the intercept and the relative weights (betas) of the variables reported by van Loendersloot and colleagues, the refitted model uses the intercept and the relative weights (betas) of the variables that were generated by a logistic regression analysis on our clinic's data. The adapted van Loendersloot model calculates IVF prognoses per complete IVF cycle at the time of the fresh embryo transfer and derives four of the five included IVF laboratory variables from the current cycle. Only the cycle variable 'embryo after oocyte retrieval (yes/no)' is based on the previous cycle. For fitting the adapted van Loendersloot model on our clinic's data, cycles without fertilisation or without an embryo suitable for transfer were excluded as these rule out the possibility of a live birth and the need for a prognosis.

The performance of all models was expressed in terms of discrimination and calibration. Discrimination refers to the ability of a model to discriminate between couples with and without a live birth after IVF. Discrimination was evaluated with the c-statistic or area under the receiver operating characteristic curve, using 'IVF prognosis' as test variable and 'observed live birth' as binary outcome.^{7 14 34-36} C-statistics range between 0.5 and 1, with 0.5 being equal to chance, higher values indicating better discrimination and 1 indicating perfect discrimination. For internal validation of the novel refitted and adapted model, leave-one-out cross-validation calculated an optimism-adjusted c-statistics with 95% CIs.

Calibration refers to the agreement between the IVF prognoses and the observed live births and was assessed

in two ways. First, we fitted a calibration model using a logistic regression with as single variable the linear combination of the variables in the model.^{9 10} Second, we compared five disjoint groups formed by the quintiles of the IVF prognoses. We described each disjoint group's ratio of the mean of the IVF prognoses and the observed live births (ie, 'prognosis to observed ratio').^{2 8} In addition, we plotted each disjoint group's mean IVF prognoses (x-axis) against their mean observed live births (y-axis).^{7 35-37} The calibration plot includes cross-points of the mean prognoses and observed live births for each disjoint group and bars representing the 95% CIs of the mean observed live births (ie, calculated with a binominal distribution model) were added. In case of perfect calibration, the five cross-points would be on the diagonal,³⁵ the five 95% CI bars would overlap with the diagonal⁹ and the 95% CI bars of the groups next to each other would not overlap.³⁵

Patient and public involvement

The research question of this study was triggered by a patient interview study concluding that patients want personalised information, rather than being given average success rates.^{3 12} Patients nor members of the public were involved in the external validation of a prognostic IVF model and in our development of two new prognostic IVF models and their validation using resampling, which were all performed on the patient data of a retrospective cohort.

RESULTS

Couples and cycles

The sample of this study includes 1281 IVF cycles from the 591 couples for whom all required clinical variables could be extracted from their EMR (ie, no missing data). Another 423 couples treated during the study period were considered 'not eligible' as at least one of the following clinical variables could not be extracted reliably from their EMR: basal FSH (n=288/423), diminished ovarian reserve (n=260/423) and duration of infertility (n=85/423). All the required IVF laboratory characteristics of the 1281 IVF cycles with a day 3 or 5 transfer performed by the 591 included couples during the study period (and for the previous cycle) could be extracted. This means there were no missing data except for the IVF laboratory variables which cannot be available if no fertilisation occurred.

The distribution of the clinical and IVF laboratory variables among the cohort of 1281 included cycles is described in [table 2](#). Online supplemental data 1 describes additional characteristics of the 1281 included IVF cycles that are not required for calculating the van Loendersloot prognosis.

The original van Loendersloot model with previous cycle data

The intercept and the relative weights (betas) of the variables reported by van Loendersloot and colleagues were

Table 2 The distribution of the clinical and IVF laboratory variables among the 1281 included cycles

Clinical variables included in both models	
Mean age of the women (SD)*	34.4±4.5
Duration of infertility (years)*	
Mean (SD)	3.4±2.2
≥5	225 (17.6%)
Previous delivery (n, %)*	432 (33.7%)
Male infertility (n, %)	692 (54.0%)
Diminished ovarian reserve (n, %)	197 (15.4%)
Endometriosis	308 (24.0%)
Basal FSH (IU/mL)	
Mean (SD)	7.2±2.5
≤10 (n, %)	1130 (88.2%)
Number of previous failed IVF/ICSI cycles (n, %)*	
0 failed IVF/ICSI cycle (n, %)	72 (5.6%)
One failed IVF/ICSI cycle (n, %)	579 (45.2%)
Two failed IVF/ICSI cycles (n, %)	340 (26.5%)
Three failed IVF/ICSI cycles (n, %)	173 (13.5%)
Four failed IVF/ICSI cycles (n, %)	90 (7.0%)
Five failed IVF/ICSI cycles (n, %)	21 (2.1%)
IVF laboratory variables included in all three models	
Fertilisation in the previous cycle (n, %)	1173 (91.6%)
IVF laboratory variables of previous cycle included in the original and refitted van Loendersloot model‡	
Number of embryos after oocyte retrieval	
Mean (SD)	5.2 (3.0)
≥10 (n, %)	105 (8.2%)
Mean of MMS of all embryos day 3 (SD)†	2.7 (0.9)
At least one 8-cell embryo on day 3 (n, %)	649/1129 (57.5%)
At least one morula on day 3 (n, %)	16/1129 (1.4%)
IVF laboratory variables of current cycle included in the adapted van Loendersloot model	
Number of embryos after oocyte retrieval	
Mean (SD)	5.3 (3.6)
≥10 (n, %)	110 (8.6%)
Mean of MMS all embryos day 3 (SD)	2.7 (0.9)
At least one 8-cell embryo on day 3 (n, %)	768/1217 (63.1%)
At least one morula on day 3 (n, %)	24/1217 (2.0%)

*Couple characteristics that vary from cycle to cycle.

†Mean MS ranging from 1 to 4 (the lower, the better).

‡These IVF laboratory variables were missing for 108/1281

cycles, as there was no fertilisation in the previous cycle.

FSH, follicle stimulating hormone; ICSI, intracytoplasmic sperm injection; IVF, in vitro fertilisation; MMS, mean morphological score.

used for this external validation.³ The calculated prognoses ranged between 0.25% and 63.14% with a mean of 14.71%.

The discrimination of the original model between couples with and without a live birth was rather poor as

the c-statistic was 0.64 (95% CI: 0.61 to 0.68). The calibration model shows an underestimation of the IVF prognosis with an estimated intercept of 0.196 (95% CI: -0.135 to 0.527) and an estimated slope of 0.643 (95% CI: 0.471 to 0.815). [Table 3](#) describes the five disjoint groups formed by the quintiles of the IVF prognoses calculated with the original model. The five disjoint groups had a predicted to observed ratio varying from 0.31 to 0.89. The calibration plot of the original model, shown in [figure 1A](#), further confirms the poor calibration. The mean IVF prognoses and the mean observed live births do not correspond well as the five disjoint groups' cross-points are far above the diagonal. None of the five disjoint groups' bars representing the 95% CIs of the mean observed live births overlap with the diagonal and the bars of the groups next to each other overlap on all four occasions.

The refitted van Loendersloot model with previous cycle data

The intercept and the relative weights of the variables of the refitted model are described in online supplemental data 2. The calculated prognoses ranged between 0.81% and 62.56% with a mean of 26.78%. The discrimination of the refitted model between couples with and without a live birth was reasonable as the c-statistic was 0.68 (95% CI: 0.65 to 0.71). Internal validation, using leave-one-out cross-validation obtained an optimistic-adjusted c-statistic of 0.65 (95% CI: 0.62 to 0.69). The calibration model confirms good calibration with an estimated intercept of 0.00 (95% CI: -0.23 to 0.23) and an estimated slope of 1.00 (95% CI: 0.79 to 1.21). [Table 4](#) describes the five disjoint groups formed by the quintiles of the IVF prognoses calculated with the refitted model. The five disjoint groups had a predicted to observed ratio varying from 0.9 to 1.0. The calibration plot of the refitted model, shown in [figure 1B](#), further confirms respectable calibration. The mean IVF prognoses and the mean observed live births correspond well as the five disjoint groups' cross-points are on or close to the diagonal. All five disjoint groups' bars representing the 95% CIs of the mean observed live births overlap with the diagonal. Unfortunately, the bars of the groups next to each other do overlap on all four occasions.

The adapted van Loendersloot model with current cycle data

The intercept and the relative weights of the variables of the adapted model are described in online supplemental data 2. The calculated prognoses ranged between 0.89% and 86.23% with a mean of 28.92%.

The discrimination of the adapted model between couples with and without a live birth was good as the c-statistic was 0.74 (95% CI: 0.70 to 0.77). Internal validation, using leave-one-out cross-validation obtained an optimistic-adjusted c-statistic of 0.71 (95% CI: 0.68 to 0.75). The calibration model confirms good calibration with an estimated intercept of 0.00 (95% CI: -0.18 to 0.18) and an estimate slope of 1.00 (95% CI: 0.83 to 1.17). [Table 5](#) describes the five disjoint groups formed by the quintiles of the IVF prognoses calculated with the

Table 3 Description of the five disjoint groups formed by the quintiles of the IVF prognoses calculated by the original van Loendersloot model (n=1.281)

Disjoint group	Range of prognoses (%)	Number of couples	Mean prognosis (%)	Number of live births	Mean observed live births (%; 95% CI)	Prognosis to observed ratio
1	0–6.5	253	4.04	33	13.04 (9.2 to 17.8)	0.31
2	6.5–11.5	267	9.12	57	21.35 (16.6 to 26.8)	0.43
3	11.5–16	243	13.82	65	26.75 (21.3 to 32.8)	0.89
4	16–21.5	260	18.46	78	30.0 (24.5 to 36.0)	0.62
5	>21.5	258	28.02	110	53.1 (36.5 to 48.9)	0.53

IVF, in vitro fertilisation.

adapted model. The predicted to observed ratio of the five disjoint groups varied between 0.9 and 1.1. The calibration plot of the adapted model, shown in [figure 1C](#), confirms very good calibration. The mean IVF prognoses and the mean observed live births correspond well as the five disjoint groups' cross-points are on or close to the diagonal. All five disjoint groups' bars representing the 95% CIs of the mean observed live births overlap with the diagonal. Only the bars of the third and fourth disjoint group overlap, while there is no overlap on the remaining three occasions.

Online supplemental data 3 shows a side-by-side comparison of the discrimination capacity of the original, the refitted and the adapted van Loendersloot model.

DISCUSSION

External validation of the original van Loendersloot prognostic model showed a poor discrimination and calibration, indicating that model updating was recommended. After updating, this study confirms the good discrimination and calibration of the previously introduced prognostic model.³⁹ Building on the work of van Loendersloot and colleagues, by replacing previous by current cycle IVF

laboratory variables, led to a very performant model with a c-statistic of 0.74 (0.71 after cross-validation) and a calibration plot practically coinciding with the diagonal.^{15 36}

The methodology of this study has several strengths. We selected a model of which the high-quality development was thoroughly reported (as reviewed by Ratna *et al*)⁶ and took account of the TRIPOD criteria to ensure high-quality methodology and a thoroughly report on the model's external validation and on model updating and validation through resampling. Data extraction was not conducted in a blinded manner, as advised by the TRIPOD statement,¹⁶ as the predictors and outcome were extracted simultaneously from the EMR. While using the van Loendersloot model, we learnt we could contribute to the transferability of the model by reporting a clear definition for all variables.¹⁴ Furthermore, internal validation using leave-one-out cross-validation, was used to obtain optimism-adjusted c-statistics for the novel refitted and adapted models. Finally, calibration was assessed with multiple methods including a calibration model and assessing the match between the prognoses and observed success rates for disjoint groups.¹⁴ Like many other studies,³⁸ the used methodology, however, had

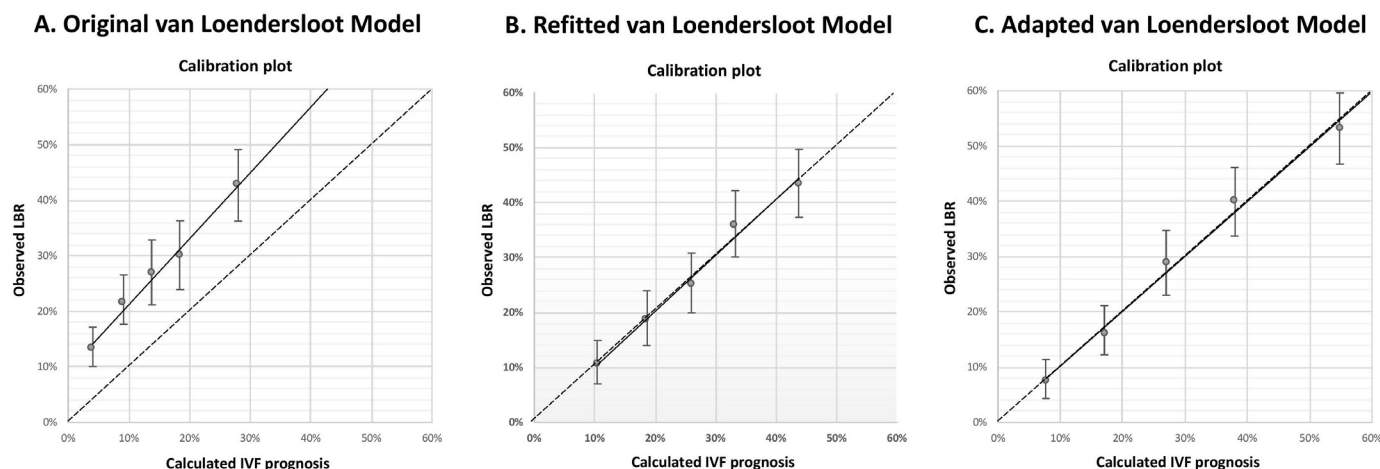


Figure 1 Side-by-side comparison of the calibration plots of the original (A), the refitted (B) and the adapted (C) van Loendersloot model. In each calibration plot, showing the relationship between the calculated IVF prognosis and observed LBRs, the five groups represent the quintiles of the calculated IVF prognoses. Data on observed LBR are reported as percentage and 95% CI. IVF, in vitro fertilisation; LBR, live birth rate.

Table 4 Description of the five disjoint groups formed by the quintiles of the IVF prognoses calculated by the refitted van Loendersloot model (n=1.281)

Disjoint group	Range of prognoses (%)	Number of couples	Mean prognosis (%)	Number of live births	Mean observed live births (%; 95% CI)	Prognosis to observed ratio
1	0–15	257	10.5	27	10.5 (7.00 to 14.9)	1.0
2	15–22.5	252	18.6	47	18.7 (14.0 to 24.0)	1.0
3	22.5–30	259	26.0	65	25.1 (19.9 to 30.8)	1.0
4	30–38	253	33.2	91	36.0 (30.1 to 42.2)	0.9
5	>38	260	43.8	113	43.5 (37.3 to 49.7)	1.0

IVF, in vitro fertilisation.

the limitation of relying on a retrospective cohort, while prospective recruitment could result in a more homologue IVF protocol and in less missing data.¹⁴ During the 8-year study period, the number of embryos per transfer was restricted by the same Belgian law. Slow freezing was replaced by vitrification, but this did not significantly alter this cohort's LBR per completed cycle ($p=0.06$). Another limitation is that we selected a sample of couples and cycles for which all clinical and IVF laboratory variables had been registered (ie, complete case analysis) and report on recruitment rates rather than performing imputation for the missing data.^{10 14} Focussing on live birth instead of ongoing pregnancy as outcome of the van Loendersloot model is in line with: (1) the previous external validation,⁹ (2) with birth rates being more relevant for patients than clinical pregnancy rates, (3) with van Loendersloot and colleagues themselves considering live birth a more ideal outcome and not expecting this outcome to fundamentally change the model³ and (4) with the external validation study reporting that using ongoing pregnancy rather than live birth did not alter the performance of the model.⁹ The transferability of the van Loendersloot model after refitting it, as appropriate for each IVF clinic wanting to use a prognostic model,^{39 40} is confirmed by the current study. The differences in the relative weights of the variables when refitting the van Loendersloot model for different settings^{3 9} were probably due to interinstitutional differences in IVF protocols, IVF success rates, definitions of variables and patient population (eg, woman's age, endometriosis)⁴¹ rather than due to replacing the outcome ongoing pregnancy

by live birth per completed cycle.⁹ As outlined in our methods, we validated the refitted and the adapted model in a cohort of 2nd–6th IVF cycles, based on the planned clinical applications of the van Loendersloot model in our clinic. Other clinics considering other clinical applications, might include first IVF cycles in their external validation cohort. The development study of the original van Loendersloot model also included first IVF cycles and found that adding cycle number as predictor had no significant additional effect nor did it result in additional significant interactions.³ Regarding discrimination, our refit confirmed the reasonable c-statistic of 0.68 (0.65 after cross-validation) identified during the development and internal validation of the van Loendersloot model,³ while another clinic's refit resulted in a lower c-statistic of 0.64.⁹ The identified c-statistic of 0.74 (0.71 after cross validation) for the Adapted van Loendersloot model by far exceeds in 2009 expected maximum c-statistic for IVF models of 0.62.⁷ Besides refitting all predictor regression coefficients, this strong discrimination seems mainly due to adapting an existing model by including optimal variables and due to including IVF laboratory variables. Two groups significantly improved the c-statistic of the original Templeton model for large (national or clinic) samples (n=1 44018; n=12901) by adding optimal variables.^{2 39} The recent post-treatment McLernon model, also including IVF laboratory variables, achieved c-statistics of 0.71–0.72 for large (multiclinic or national) cohorts (n=1 13 873 and n=1511).^{5 10}

Regarding calibration, the identified exceptional calibration of the adapted van Loendersloot model can partly

Table 5 Description of the five disjoint groups formed by the quintiles of the IVF prognoses calculated by the adapted van Loendersloot model (n=1.186)

Disjoint group	Range of prognoses (%)	Number of couples	Mean prognosis (%)	Number of live births	Mean observed live births (%; 95% CI)	Prognosis to observed ratio
1	0–12	234	7.7	17	7.3 (4.3 to 11.4)	1.1
2	12–22	246	17.1	39	15.9 (11.5 to 21.0)	1.1
3	22–32	234	27.0	67	28.6 (22.9 to 34.9)	0.9
4	32–45	231	38.0	92	39.8 (33.5 to 46.5)	1.0
5	>45	241	54.9	128	53.1 (46.6 to 59.5)	1.0

IVF, in vitro fertilisation.



be explained by the exclusion of cycles without embryo transfer for this model and by including IVF laboratory variables of the current cycle. An older model with less optimal but respectable performance already included current cycle IVF laboratory variables.⁴² Predicting success at the time of embryo transfer was for a long time considered of little clinical importance^{42 43} but professionals have recently acknowledged that prognostic models might not only impact (clinician's) selection of couples prior to IVF but might also be of benefit for shaping couple's IVF expectations during the procedure.⁵

The reported range between the minimal and maximal prognosis of couples from our sample (0.81% to 62.56% and 0.89% to 86.23%) demonstrates the relevance of giving couples personalised prognoses rather than using average (national or clinic) LBRs. The refitted and the adapted van Loendersloot model can now be used in our clinic to inform couples on their personalised prognosis before starting and during each cycle. In this manuscript, we report on the cross-validation of the two novel models and their internal validity will be re-evaluated on a prospectively collected cohort, which we are currently recruiting.⁴ Other clinics are encouraged to evaluate the performance of both novel models on their own data (ie, external validation).^{15 16}

The clinical applications of the refitted and the adapted van Loendersloot models are complementary to those of the more recent pretreatment and post-treatment McLernon models for which external validity was also recently proven.^{5 10} The pretreatment McLernon model is useful for helping couples and their clinicians decide whether or not to embark on an IVF journey, constituting of a package of several IVF cycles. The post-treatment McLernon model is useful for explaining couples after their first IVF cycle whether this first cycle changed what to expect from their complete IVF journey. Both McLernon models can be applied for explaining couples that they will need to engage with several IVF cycles as they calculate the success rate of a package of one to six IVF cycle(s) including the first cycle.^{5 10} Compared with the McLernon models, the van Loendersloot models provide a shorter-term perspective, as they calculate success rates per complete cycle (ie, including all fresh and frozen embryo transfers from one episode of ovarian stimulation). The van Loendersloot models can, however, be used in between cycles throughout the entire IVF journey rather than only prior to or right after embarking on the IVF journey.^{9 11} More specifically, the van Loendersloot models enable explaining couples prior to (the refitted version) or during their 2nd–6th IVF cycle (the adapted version), which success rate to expect from one complete IVF cycle. Clinicians wanting to provide an even shorter-term perspective, could inform couples on success rates per fresh embryo transfer with the interesting model of Nelson and Lawlor.^{2 4 8 44}

To our knowledge, the impact of the clinical application of IVF prediction models has yet to be studied. It would be highly interesting to evaluate whether giving a

personalised prognosis truly affects couple's IVF expectations and whether this in turn influences couple's distress and treatment decisions. Aiming to study couple's IVF expectations after having given a personalised prognosis is relevant as the general population overestimates the success of IVF^{45 46} and as women cryopreserving their oocytes think that average success rates do not apply to them.¹³ Aiming to study whether giving personalised prognoses causes distress in couples is relevant as personalised prognoses are often lower than average (national or clinic) prognoses and can be as low as 1%. Finally, aiming to study the effect of giving a personalised prognosis on treatment decisions is relevant as women shared in interviews that the combination of unrealistically high expectations and repeated unsuccessful IVF cycles led to their decision to discontinue IVF.⁴⁷

Author affiliations

¹Leuven University Fertility Centre, University Hospital Leuven, Leuven, Belgium

²Development and Regeneration, Laboratory of Endometrium, Endometriosis & Reproductive Medicine, Leuven, Belgium

³Public Health and Primary Care, Leuven Biostatistics and statistical Bioinformatics Centre, Leuven, Belgium

⁴Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Leuven, Belgium

⁵Postdoctoral fellow, Research Foundation, Flanders, Belgium

Contributors JD, KP, JV and ED contributed substantially to the conception and design of this study. JD, KP and CS made an essential effort for the acquisition of the data. JD, GV and ED analysed and interpreted the data. JD and ED prepared this manuscript and all authors read, revised and approved the final manuscript.

Funding We are very grateful to the support from the Research Council of the KU Leuven (OT/13/113). EAFD holds a Postdoctoral fellowship of the Research Foundation of Flanders.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study has been approved by the Medical Ethical Committee of the University Hospitals Leuven (Belgium).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request: Johanna.devroe@uzleuven.be.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Johanna Devroe <http://orcid.org/0000-0002-4698-9438>

Eline Dancet <http://orcid.org/0000-0003-0356-930X>

REFERENCES

- 1 Templeton A, Morris JK, Parslow W. Factors that affect outcome of in-vitro fertilisation treatment. *Lancet* 1996;348:1402–6.
- 2 Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: a prospective study of 144,018 treatment cycles. *PLoS Med* 2011;8:e1000386.
- 3 van Loendersloot LL, van Wely M, Repping S, et al. Individualized decision-making in IVF: calculating the chances of pregnancy. *Hum Reprod* 2013;28:2972–80.
- 4 Smith ADAC, Tilling K, Lawlor DA, et al. External validation and calibration of IVFPredict: a national prospective cohort study of 130,960 in vitro fertilisation cycles. *PLoS One* 2015;10:e0121357.

- 5 McLernon DJ, Steyerberg EW, Te Velde ER, *et al.* Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113 873 women. *BMJ* 2016;355:i5735.
- 6 Ratna MB, Bhattacharya S, Abdulrahim B, *et al.* A systematic review of the quality of clinical prediction models in in vitro fertilisation. *Hum Reprod* 2020;35:100–16.
- 7 Coppus SFPJ, van der Veen F, Opmeer BC, *et al.* Evaluating prediction models in reproductive medicine. *Hum Reprod* 2009;24:1774–8.
- 8 te Velde ER, Nieboer D, Lintsen AM, *et al.* Comparison of two models predicting IVF success; the effect of time trends on model performance. *Hum Reprod* 2014;29:57–64.
- 9 Sarais V, Reschini M, Busnelli A, *et al.* Predicting the success of IVF: external validation of the Van Loendersloot's model. *Hum Reprod* 2016;31:1245–52.
- 10 Leijdekkers JA, Eijkemans MJC, van Tilborg TC, *et al.* Predicting the cumulative chance of live birth over multiple complete cycles of in vitro fertilization: an external validation study. *Hum Reprod* 2018;33:1684–95.
- 11 van Loendersloot LL, van Wely M, Repping S, *et al.* Templeton prediction model underestimates IVF success in an external validation. *Reprod Biomed Online* 2011;22:597–602.
- 12 Dancet EAF, Van Empel IWH, Rober P, *et al.* Patient-centred infertility care: a qualitative study to listen to the patient's voice. *Hum Reprod* 2011;26:827–33.
- 13 de Groot M, Dancet E, Repping S, *et al.* Perceptions of oocyte banking from women intending to circumvent age-related fertility decline. *Acta Obstet Gynecol Scand* 2016;95:1396–401.
- 14 Bouwmeester W, Zuihoff NPA, Mallett S, *et al.* Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012;9:1–12.
- 15 Moons KGM, Kengne AP, Grobbee DE, *et al.* Risk prediction models: II. external validation, model updating, and impact assessment. *Heart* 2012;98:691–8.
- 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;162:W1–73.
- 17 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55–63.
- 18 Mansour R, Ishihara O, Adamson GD, *et al.* International Committee for monitoring assisted reproductive technologies world report: assisted reproductive technology 2006. *Hum Reprod* 2014;29:1536–51.
- 19 Zegers-Hochschild F, Adamson GD, Dyer S, *et al.* The International glossary on infertility and fertility care, 2017. *Fertil Steril* 2017;108:393–406.
- 20 European IVF-monitoring Consortium (EIM), European Society of Human Reproduction and Embryology (ESHRE), Calhaz-Jorge C, *et al.* Assisted reproductive technology in Europe, 2013: results generated from European registers by ESHRE. *Hum Reprod* 2017;32:1957–73.
- 21 Vergouwe Y, Steyerberg EW, Eijkemans MJC, *et al.* Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;58:475–83.
- 22 Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214–26.
- 23 Debrock S, Melotte C, Spiessens C, *et al.* 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.
- 24 Debrock S, Peeraer K, Fernandez Gallardo E, *et al.* Vitrification of cleavage stage day 3 embryos results in higher live birth rates than conventional slow freezing: a RCT. *Hum Reprod* 2015;30:1820–30.
- 25 Paternot G, Debrock S, De Neubourg D, *et al.* Semi-automated morphometric analysis of human embryos can reveal correlations between total embryo volume and clinical pregnancy. *Hum Reprod* 2013;28:627–33.
- 26 De Neubourg D, Bogaerts K, Wyns C, *et al.* The history of Belgian assisted reproduction technology cycle registration and control: a case study in reducing the incidence of multiple pregnancy. *Hum Reprod* 2013;28:2709–19.
- 27 IVF. *Modalities reimbursement IVF lab activities*, 2003.
- 28 Paternot G, Debrock S, D'Hooghe TM, *et al.* Early embryo development in a sequential versus single medium: a randomized study. *Reprod Biol Endocrinol* 2010;8:83.
- 29 Fernandez Gallardo E, Spiessens C, D'Hooghe T, *et al.* Effect of day 3 embryo morphometrics and morphokinetics on survival and implantation after slow freezing-thawing and after vitrification-warming: a retrospective cohort study. *Reprod Biol Endocrinol* 2017;15:79.
- 30 Sociale Zekerheid. *Wet betreffende de medisch begeleide voortplanting en de bestemming van de overtallige embryo's en de gameten*, 2007: 17.
- 31 Debrock S, Peeraer K, Spiessens C, *et al.* The effect of modified quarter laser-assisted zona thinning on the implantation rate per embryo in frozen/vitrified-thawed/warmed embryo transfer cycles: a prospective randomized controlled trial. *Hum Reprod* 2011;26:1997–2007.
- 32 WHO. *World Health Organization laboratory manual for the examination and processing of human semen*. Geneva, Switzerland: World Health Organization, 2010.
- 33 Ferraretti AP, La Marca A, Fauser BCJM, *et al.* ESHRE consensus on the definition of 'poor response' to ovarian stimulation for in vitro fertilization: the Bologna criteria. *Hum Reprod* 2011;26:1616–24.
- 34 Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem* 2008;54:17–23.
- 35 Leushuis E, van der Steeg JW, Steures P, *et al.* Prediction models in reproductive medicine: a critical appraisal. *Hum Reprod Update* 2009;15:537–52.
- 36 van Loendersloot L, Repping S, Bossuyt PMM, *et al.* Prediction models in in vitro fertilization; where are we? A mini review. *J Adv Res* 2014;5:295–301.
- 37 Cox DR. The regression analysis of binary sequences. *J R Stat Soc Series B* 1958;20:215–32.
- 38 Zarinara A, Zeraati H, Kamali K, *et al.* Models predicting success of infertility treatment: a systematic review. *J Reprod Infertil* 2016;17:68–81.
- 39 Arvis P, Leher P, Guivarc'h-Levêque A. Simple adaptations to the Templeton model for IVF outcome prediction make it current and clinically useful. *Hum Reprod* 2012;27:2971–8.
- 40 Van Calster B, McLernon DJ, van Smeden M, *et al.* Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;17:230.
- 41 Lintsen AME, Braat DDM, Habbema JDF, *et al.* Can differences in IVF success rates between centres be explained by patient characteristics and sample size? *Hum Reprod* 2010;25:110–7.
- 42 Stolwijk AM, Zielhuis GA, Hamilton CJ, *et al.* Prognostic models for the probability of achieving an ongoing pregnancy after in-vitro fertilization and the importance of testing their predictive value. *Hum Reprod* 1996;11:2298–303.
- 43 Mol BWJ, van Wely M, Steyerberg EW. Using prognostic models in clinical infertility. *Hum Fertil* 2000;3:199–202.
- 44 Saha L, Fook-Chong SM, Rajesh H, *et al.* Use of in vitro fertilisation prediction model in an Asian Population-Experience in Singapore. *Ann Acad Med Singap* 2015;44:524–9.
- 45 Lampic C, Svanberg AS, Karlström P, *et al.* Fertility awareness, intentions concerning childbearing, and attitudes towards parenthood among female and male academics. *Hum Reprod* 2006;21:558–64.
- 46 Skoog Svanberg A, Lampic C, Karlström P-O, *et al.* Attitudes toward parenthood and awareness of fertility among postgraduate students in Sweden. *Gend Med* 2006;3:187–95.
- 47 Peddie VL, van Teijlingen E, Bhattacharya S. A qualitative study of women's decision-making at the end of IVF treatment. *Hum Reprod* 2005;20:1944–51.