RBM ONLINE

ELSEVIER

Article

Predicting live birth for poor ovarian responders: the PROsPeR concept

Philippe Lehert [a,b,*], ***Wai Chin*** [c,1], ***Joan Schertz*** [c,1], ***Thomas D'Hooghe*** [d,e], ***Carlo Alviggi*** [f,g], ***Peter Humaidan*** [h,i]

[a] *Faculty of Economics, UCL Mons, Chaussée de Binche 151, Mons 7000, Belgium*
[b] *Faculty of Medicine, The University of Melbourne, Melbourne, Australia*
[c] *EMD Serono Research and Development Institute, 45A Middlesex Turnpike, Billerica, MA 01821-3936, USA*
[d] *Merck KGaA, Frankfurter Str. 250, 64293, Darmstadt, Germany*
[e] *Department of Development and Regeneration, University of Leuven (KU Leuven), Leuven, Belgium*
[f] *Dipartimento di Scienze Anatomiche, Istologiche, Medico-Legali e dell'Apparato Locomotore, Sezione Istologia ed Embriologia Medica, La Sapienza, University of Rome, Rome, Italy*
[g] *Dipartimento di Neuroscienze, Scienze Riproduttive ed Odontostomatologiche, University Federico II, via Sergio Pansini 5, 80131, Naples, Italy*
[h] *The Fertility Clinic, Skive Regional Hospital, Resenvej 25, 7800 Skive, Denmark*
[i] *Faculty of Health, Aarhus University, Aarhus, Denmark*

Philippe Lehert is Professor of Mathematical Statistics and Biostatistics at the Faculty of Medicine, University of Melbourne (Australia) and at the Faculty of Economics, University of Louvain (Belgium). He was a senior adviser for the United Nations (1994–2011), and has been a member of the Cochrane Review Group since 1996.

KEY MESSAGE
PROsPeR is a simple and accurate live-birth estimate designed specifically for poor ovarian responders and is suitable for routine practice. PROsPeR produces a score (0, 1 or 2), with the predicted live-birth rate reduced by a factor of three with each additional predictor present.

A B S T R A C T

Research Question: A number of live-birth predictive models are available, and despite clinical interest these are rarely used owing to poor performance. In addition, no predictive models specifically for poor ovarian responders (POR) are available. The aim of the current project was to develop a clinically applicable tool for predicting live birth for PORs receiving recombinant human FSH [r-hFSH].

Design: A model was developed to predict live birth in PORs receiving r-hFSH, using data from the ESPART trial. Initially, two models were developed separately: one for patients with data from a previous assisted reproductive technology (ART) cycle and one for ART treatment-naïve patients. Subsequently, the simplified Poor Responder Outcome Prediction (PROsPeR) concept was derived.

* Corresponding author.
E-mail address: philippe.lehert@gmail.com (P Lehert).
[1] A business of Merck KGaA, Darmstadt, Germany.

https://doi.org/10.1016/j.rbmo.2018.03.013

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013

Results: PROsPeR considers three predictors and categorizes PORs into three scores, with predicted the live-birth rate divided by three with each worsening category. When adequately calibrated, a discrimination score up to area under the receiver operating characteristic (AUC_{ROC}) (95% CI) of 0.84 (0.79 to 0.88) was observed, which is superior to previously published models. Lower discriminations were observed when the PROsPeR model was used to evaluate the patients who received both r-hFSH and recombinant human LH in the ESPART study (AUC_{ROC} [95% CI] 0.66 [0.61 to 0.71]) and when all the patients included in the ESPART study were evaluated (AUC_{ROC} [95% CI] 0.68 [0.61 to 0.72]).

Conclusions: This model, specific to PORs receiving r-hFSH, constitutes the best compromise between precision and simplicity, and is suitable for routine practice.

Introduction

Assisted reproductive technology (ART) treatment can have a significant emotional effect on patients, making appropriate counselling of patients an essential aspect of infertility care (Verhaak et al., 2010; American Society for Reproductive Medicine, 2016; Eijkemans et al., 2017). This includes the management of expectations before treatment is initiated, in particular ensuring that patients are aware of their chances of success. Furthermore, to ensure patients have the greatest chance of a live birth, it is important that treatment is individualized, including selection of the most appropriate treatment and dosing for ovarian stimulation (La Marca and Sunkara, 2014a, 2014b; Teixeira and Martins, 2014). A number of models have been developed with the aim of predicting outcomes for the infertile patient. These might also be used to individualize treatment and may be relevant when developing clinical trials (McLernon et al., 2016; Nelson and Lawlor, 2011; Porcu et al., 2013; Templeton et al., 1996).

The best known of these, the Templeton, Nelson and McLernon models, were developed using data from the UK Human Fertilisation and Embryology Authority. The Templeton model, which was created using data from 39,961 cycles carried out between 1991 and 1994, is generally considered the best despite poor discrimination, which is the extent to which the prediction is exact (Templeton et al., 1996). Furthermore, it is the only model to have been externally validated using non-UK data (Leushuis et al., 2009). The Nelson model was developed more recently using data from 144,018 cycles carried out between 2003 and 2007 (Nelson and Lawlor, 2011), and has been externally validated using UK data, demonstrating comparable performance to the Templeton model (Smith et al., 2015). More recently, McLernon et al. (2016) developed two complementary models using data from 184,269 fresh and frozen–thawed embryo transfer cycles carried out between 1999 and 2008, one based on predictors available before treatment is initiated and the other including predictors available once treatment has been commenced (McLernon et al., 2016). The latter two models still require external validation. These models involve similar, well-established predictors, including, but not limited to, female age, duration of infertility, number of previous successful or unsuccessful IVF cycles (an IVF cycle is considered successful if it results in a live birth), pregnancy history and whether infertility was caused by tubal pathology.

Despite interest in predicting outcomes, these models are rarely used in clinical practice, as their discrimination and calibration remain unconvincing (Arvis et al., 2012; Leushuis et al., 2009; van Loendersloot et al., 2011). This lack of precision has several potential explanations: (A) a model aiming to predict outcomes for all patients may be too ambitious, as predictors may affect outcomes differently depending on the cause and severity of infertility or on the particular subgroup (La Marca and Sunkara, 2014a); (B) within these patient subgroups, important effects of the type and dose of medication selected for ovarian stimulation were found (Lehert et al., 2014; Mochtar et al., 2017; Santi et al., 2017). Furthermore, models not accounting for a given medication fail to provide comparisons between alternative treatments, which probably constitutes the most important use of these models in actual clinical practice; (C) the centre has a major effect on live-birth prediction, and the wholesale application of a predictive model whose development was based on data from one or several centres to another centre is known to provide biased predictions (Arvis et al., 2012); (D) the models are not easy to use, requiring a computer to conduct the calculations.

Despite knowledge of these issues, to date, no practical solution has been proposed for easily adapting a simple predictive model that accounts for a specific treatment, subpopulations and centres.

For the model described herein, a pragmatic decision was made about the choice of patient subgroup, based on the availability of data, and considerations about which subgroup would benefit most from a predictive model. Poor ovarian response (POR), which affects between 5.6% and 35.1% of the infertile population (Oudendijk et al., 2012), is associated with the poorest reproductive outcomes (La Marca et al., 2015; Yang et al., 2016). Counselling and appropriate treatment selection are, therefore, of particular importance for these patients. To date, models have been developed using data from the overall population of infertile patients, and their applicability for patients with POR remains unproven. Therefore, the development of such a tool is not only justified, but can be considered to be a priority.

Owing to the fairly recent release of the European Society of Human Reproduction and Embryology Bologna Criteria (EBC) in 2011 (Ferraretti et al., 2011), no specific large datasets are available to build predictive models for patients with POR and, because of the likelihood of missing data, a retrospective database analysis would likely provide lower quality data with which to develop such a model. The ESPART (Efficacy and Safety of Pergoveris® in Assisted Reproductive Technology) study, the largest randomized controlled trial in patients with POR to date, investigated the effect of recombinant human LH (r-hLH) supplementation to recombinant human FSH (r-hFSH) for ovarian stimulation in patients with POR using criteria based on, but stricter than, the EBC (Humaidan et al., 2017a). These stricter criteria, including exclusion of women aged 41 years or over, were used to remove diagnostic subjectivity, reduce patient heterogeneity and exclude patients with the worst reproductive prognosis (Humaidan et al., 2017a). Moreover, an extensive list of baseline variables was recorded during the ESPART study that could be used as outcome

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013

predictors. Data from this study provided the first opportunity to develop an innovative predictive model of live birth for patients with POR that might be relevant to everyday clinical practice.

The ESPART study compared r-hFSH alone with fixed-ratio (2:1) r-hFSH and r-hLH for ovarian stimulation. As r-hFSH is the reference product for ovarian stimulation, and has been used for more than 20 years, the decision was taken to base the model development on data from patients receiving r-hFSH alone.

Therefore, the aim of the current project was to develop a clinically applicable tool for predicting live birth in patients with POR, receiving a specific treatment (r-hFSH) that was accurate, simple to use and easily adaptable for a single centre.

Materials and methods

ESPART study data

To be enrolled in the ESPART study, women had to meet at least two of the following POR criteria: advanced maternal age (≥40– < 41 years); a previous ART cycle with three or fewer oocytes retrieved with a conventional stimulation protocol; an abnormal ovarian reserve test characterized by an anti-Müllerian hormone (AMH) level between 0.12 and 1.3 ng/ml, inclusive (measured by AMH GEN II ELISA, Beckman Coulter, Inc., High Wycombe, UK). Patients were down-regulated with a long gonadotrophin releasing hormone agonist protocol using daily triptorelin acetate (Decapeptyl, Ferring Pharmaceuticals, Saint-Prex, Switzerland). Once successful down-regulation was confirmed, patients received ovarian stimulation with r-hFSH (GONAL-f®, Merck KGaA, Darmstadt, Germany) 300 IU or fixed ratio (2:1) r-hFSH /r-hLH (Pergoveris®, Merck KGaA, Darmstadt, Germany) 300 IU/150 IU, with the dose maintained for the first 4 days of stimulation. Dose adjustments (either increases or decreases of 75 IU) were then permitted, with a maximum allowed daily dose of 450 IU r-hFSH (and 225 IU r-HLH in the group receiving r-hFSH /r-hLH). Once follicle(s) reached a mean diameter of 17–18 mm, a single injection of recombinant hCG (Ovidrel® Prefilled Syringe, Merck KGaA, Darmstadt, Germany) was given to trigger final follicular maturation. Oocyte retrieval occurred 34–38 h after the administration of r-hCG, and subsequent embryo transfer (maximum three embryos transferred) occurred according to each centre's standard practice 2–3 days after oocyte retrieval (Humaidan et al., 2017a). Patients were followed up until the outcome of treatment could be confirmed.

When developing the predictive models, the intention-to-treat population was used for the primary analyses, with the per-protocol population, a subset of the intention-to-treat population including all women who did not have any major protocol deviations that were likely to affect efficacy, used for sensitivity analyses. Only data from patients who received r-hFSH alone, i.e. no supplementation with r-hLH, were used during model development. The primary outcome investigated in the models was live birth, and data for both patients with or without live birth were used in their development. Validation was then carried out using data from the r-hFSH /r-hLH arm of the ESPART study as well as the entire ESPART study population, to confirm whether there was an identifiable treatment effect.

Statistical analysis

On the basis of the literature findings and expert opinion meetings, the candidate baseline predictors predefined in the statistical analysis plan to be evaluated for inclusion in the model were: age, number of oocytes obtained in a previous ART treatment cycle (PNO), AMH level, antral follicle count (AFC), infertility type (primary or secondary), cause of infertility (tubal, endometriosis, ovulatory problems or unexplained), duration of infertility, previous number of pregnancies, previous number of ART cycles, country and study centre. On the basis of a a literature search, age was evaluated as either a linear or quadratic effect (age^2), and PNO and AMH were evaluated as either linear or logarithmic effects. Missing data were imputed by maximum likelihood estimation.

The predictive model was developed using a non-linear mixed model featuring logistic regression, with live birth as a dependent binary end-point and each of the baseline characteristics as fixed covariates. Study centre nested into country effect was included as a random factor for both intercept and baseline severity variables. A shrinkage factor was included (Lasso regression) to reduce the overfit of the model and obtain relatively unbiased estimates (Steyerberg et al., 2001). A backward strategy was used to simplify the model, based on the maximization of the Akaike Information Criterion, with each elimination tested using a hierarchical F-test for fixed covariates and likelihood ratio test for mixed factors. For sensitivity purposes, the backward analysis was compared with a stepwise strategy where the introduction of a new predictor was conditional on net reclassification improvement (Pencina et al., 2010).

The results of these statistical analyses were discussed with clinicians, who requested a simpler and easier to use method of calculation. Simplification of the model was, therefore, attempted, substituting continuous variables with dichotomized variables based on pre-determined thresholds. Simplification in this manner is not recommended owing to loss of information and the potential for bias (Royston et al., 2006). Simplification was therefore conducted with care and the knowledge that non-linear effects had been observed for the included variables. Thresholds for dichotomized variables were determined from the ESPART study data, based on the maximization of sensitivity and specificity (Youden index) of each predictor in relation to the live-birth rate and checked for stability through random subsampling. These thresholds were also compared with thresolds set in other published studies and their clinical relevance discussed with clinicians.

The validity of the general model (including all the variables), initial model (including continuous significant variables) and simplified model (based on dichotomous variables) were then compared based on their discrimination (AUC_{ROC}, C-statistics) (Hanley and McNeil, 1982), both unadjusted and adjusted for centre performance. Model calibration was evaluated by the goodness of fit test of observed versus predicted frequencies for equally sampled categories (Hosmer and Lemeshow, 2000) with a correction factor (Harrell et al., 1996). The Nagelkerke coefficient of determination adapted for binary end-points was reported for each model.

For sensitivity purposes, the validity of the r-hFSH model was verified when applied to the r-hFSH /r-hLH arm alone and to the entire ESPART population, using the techniques described above.

The statistical package R (release 2.3.4; The R Foundation) was used for all statistical analyses.

Results

The baseline characteristics of patients included in this analysis are shown in **Table 1**. Of the 477 patients who received r-hFSH, 427 had

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013

Table 1 – Baseline characteristics of patients in the ESPART study.

Baseline characteristics	r-hFSH (*n* = 477)	r-hFSH/r-hLH (*n* = 462)
AMH (ng/ml)[a]	0.60 ± 0.48	0.58 ± 0.50
AFC[b]	4.8 ± 2.2	4.9 ± 2.3
Age (years)	38.3 ± 3.0	38.3 ± 2.9
PNO	2.50 ± 1.80	2.52 ± 1.96
AMH <0.5 ng/ml, n/N (%)	232/475 (48.8)	235/458 (51.3%)
Age ≥40 years, n/N (%)	236/477 (49.5)	229/462 (49.6)
PNO <2, n/N (%)	127/427 (29.7)	113/402 (28.1%)
Duration of infertility, years	4.4 ± 3.5	4.6 ± 3.7
Number of previous ART cycles	1.73 ± 1.21	1.75 ± 1.20
Number of previous live births	0.22 ± 0.47	0.20 ± 0.44

Data are mean ± SD unless otherwise stated.

[a] Data are missing for two patients who received r-hFSH and four patients who received r-hFSH/r-hLH.

[b] Data are missing for six patients who received r-hFSH and seven patients who received r-hFSH/r-hLH.

AFC, antral follicle count; AMH, anti-Müllerian hormone; ART, assisted reproductive technology; PNO, number of oocytes retrieved in a previous ART treatment cycle; r-hFSH, recombinant human FSH; r-hLH, recombinant human LH; SD, standard deviation.

data available from a previous ART treatment cycle, and 56 (11.7%) patients had a live birth. Of the 462 patients who received r-hFSH /r-hLH, 402 had data available from a previous ART treatment cycle, and 49 (10.6%) patients had a live birth. Outcome data were available for all patients. Few data were missing: AMH and AFC values were missing for two and six patients, respectively, who received r-hFSH and four and seven patients, respectively, who received r-hFSH /r-hLH.

Development (r-hFSH data only)

The strongest effect was observed for PNO, which meant that two models had to be developed using data for all patients who received r-hFSH alone (*n* = 477): one which included predictors for patients experienced with ART (*n* = 427) and one which excluded predictors available after a previous treatment cycle, thereby differentiating the outcomes for patients who were inexperienced with ART treatment.

For patients who were experienced with ART, two determinant predictors were identified after backwards analysis (**Table 2**). These were a non-linear age effect (OR 0.993/age^2; 95% CI 0.989 to 0.998; $P = 0.002$) and a linear effect from the log-transformed PNO (OR 3.596; 95% CI 1.261 to 9.256; $P = 0.017$). When only predictors that would be available for patients inexperienced with ART treatment were considered, a linear AMH effect (OR 1.724; 95% CI 1.035 to 2.872; $P = 0.037$) was observed, in addition to the same non-linear age effect included in the model for patients who were experienced with ART treatment (OR 0.994/age^2; 95% CI 0.990 to 0.998; $P = 0.004$).

As AMH and AFC were correlated, they were tested separately as continuous variables and had similar effects on live birth, although the effect of AMH was slightly superior (AMH: $P = 0.03$; AFC: $P = 0.07$). Data for models developed using AFC rather than AMH are shown in **Table 3**.

Simplification

The dichotomization of the continuous predictors was carefully assessed, based on the maximization of the sum of sensitivity and specificity (Youden index) for live birth and testing for loss of information. The optimally identified thresholds for age, AMH and PNO (**Table 2**) remained stable across random subsampling and matched literature findings (**Table 4**). Restarting backward analyses with both continuous and dichotomized predictors provided the same predictors as the initial models but favoured the dichotomized variables. The first binary condition, age 40 years or over, was found in both models, whereas a second condition, PNO less than 2 for patients experienced with ART treatment or AMH less than 0.5 ng/ml for patients inexperienced with ART treatment, reflected ovarian reserve measured by two surrogate markers. The Poor Responder Outcome Prediction (PROsPeR), reported herein, produces a score based on

Table 2 – Comparison of the initial model based on continuous predictors and the simplified model (based on dichotomized predictors).

		Patients experienced with ART (*n* = 427)			Patients inexperienced with ART[a] (*n* = 477)		
		OR	95% CI	*P*-value	OR	95% CI	*P*-value
Initial model	Intercept[b]	0.355	0.100, 1.265	NS	0.294	0.068, 1.268	NS
	Age, years (Age^2)	0.993	0.989, 0.998	0.002	0.994	0.990, 0.998	0.004
	PNO, Log(PNO+0.1)	3.596	1.261, 9.256	0.017	–	–	–
	AMH, ng/ml	1.540	0.895, 2.649	NS	1.724	1.035, 2.872	0.037
	Infertility duration, years	0.954	0.860, 1.059	NS	0.938	0.846, 1.040	NS
	Infertility type	0.936	0.392, 2.236	NS	0.881	0.381, 2.034	NS
	Number of cycles	0.966	0.745, 1.254	NS	1.070	0.846, 1.354	NS
	Number of live births	1.159	0.481, 2.788	NS	1.057	0.465, 2.402	NS
	Determination R^2	0.074	0.023, 0.123	<0.001	0.075	0.013, 0.125	<0.001
Simplified model	Intercept[b]	0.289	0.200, 0.416	<0.001	0.280	0.185, 0.423	<0.001
	Age ≥40 years	0.290	0.149, 0.565	<0.001	0.362	0.198, 0.665	0.001
	PNO <2	0.189	0.075, 0.478	0.001	–	–	–
	AMH <0.5 ng/ml	–	–	–	0.443	0.243, 0.807	0.008
	Determination R^2	0.073	0.019, 0.145	<0.001	0.069	0.021, 0.121	<0.001

The dashes indicate where data are not available as the variables were not evaluated.

[a] Data for all patients treated with r-hFSH were used, excluding predictors that would only be available for patients experienced with ART.

[b] Intercept value corresponds to the live birth estimate for a population of women defined by age 25 years, AMH 1 ng/ml, infertility duration 5 years, and two unsuccessful prior cycles.

AMH, anti-Müllerian hormone; ART, assisted reproductive technology; NS, non-significant; PNO, number of oocytes retrieved in a previous ART treatment cycle.

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013

Table 3 – Predictive model of live birth including antral follicle count rather than anti-Müllerian hormone.

	Patients experienced with ART (*n* = 427)			Patients inexperienced with ART[a] (*n* = 477)		
	OR	95% CI	*P*-value	OR	95% CI	*P*-value
Intercept[b]	0.242	0.047, 1.238	NS	0.255	0.051, 1.268	NS
Age, years (Age^2)	0.994	0.990, 0.999	0.011	0.995	0.991, 0.999	0.013
PNO (Log[PNO])	4.263	1.438, 12.633	0.009	–	–	–
AFC	0.971	0.780, 1.209	NS	0.943	0.883, 1.008	NS
AFC^2	0.946	0.876, 1.022	NS	0.953	0.863, 1.053	NS
Infertility duration, years	0.963	0.868, 1.068	NS	0.803	0.351, 1.841	NS
Infertility type	0.863	0.356, 2.090	NS	1.060	0.834, 1.347	NS
Number of cycles	0.964	0.731, 1.272	NS	1.128	0.494, 2.574	NS
Number of live births	1.232	0.505, 3.004	NS	0.255	0.051, 1.268	NS
Determination R^2	0.065	0.021, 0.133	<0.001	0.061	0.011, 0.129	<0.001

The dashes indicate where data are not available as the variable is not included in the model.

[a] Data for all patients treated with recombinant human FSH was used, excluding predictors that would only be available for patients experienced with ART treatment.

[b] The intercept value corresponds to the live birth estimate for a population of women defined by age 25 years, AMH 1 ng/mL, infertility duration 5 years, and two unsuccessful prior cycles.

ART, assisted reproductive technology; PNO, number of oocytes retrieved in a previous ART treatment cycle.

these two specified binary conditions; the score can be 0 (neither condition met), 1 (one of the two conditions met) or 2 (both conditions met). When logistic regression was conducted on live-birth rates, with the PROsPeR score as the unique categorical predictor, highly significant effects were found for a score of 1 (OR 0.29; 95% CI 0.16 to 0.53) and 2 (OR 0.17; 95% CI 0.05 to 0.56]), compared with a score of 0. Finally, the linear approximation of the score was found acceptable (OR 0.33 per score; 95% CI 0.20 to 0.55; Akaike Information Criterion decrease, and chi-squared deviance test non-significant decrease compared with the categorical model). For this last model, a highly significant random centre effect was identified for the intercept (mean live birth rate over all centres: 19.1%; 95% CI 4.2% to 73.1%), whereas the PROsPeR score effect did not significantly vary (random PROsPeR effect over centres and countries).

Model validation and comparison

The validity of the general, initial, simplified and PROsPeR predictions was assessed for centre, adjusted and unadjusted for patients treated with r-hFSH alone (**Table 5**). An acceptable discrimination score was observed for all the predictive models for r-hFSH treatment, with the lowest AUC_{ROC} (95% CI) for PROsPeR (0.709 [0.639, 0.779]). However, despite the decrease attributable to the approximation of the model, it is not significantly inferior to the others. The discrimination of PROsPeR improved to 0.835 (0.786, 0.883) when adjusted for centre effect. For each model, calibration was successfully tested by the Hosmer–Lemeshow test, and the fitted line did not significantly depart from the diagonal (slope: 0.94; 95% CI 0.84–1.59 for PROsPeR).

When PROsPeR was applied to the data for patients who received r-hFSH and r-hLH, the discrimination was lower (AUC_{ROC} [95% CI] of 0.564 [0.463, 0.662] and 0.655 [0.609, 0.711] for results unadjusted and adjusted for centre, respectively); however, calibration remained acceptable (Hosmer–Lemeshow test $P > 0.05$). Intermediate results, between those for r-hFSH and those for r-hFSH + r-hLH, were found when PROsPeR was applied to the full ESPART population (r-hFSH and r-hFSH + r-hLH), values of AUC_{ROC} (95% CI) were 0.627 (0.588 to 0.677) and 0.676 (0.613 to 0.723) for centre-unadjusted and centre-adjusted results, respectively, and calibration remained acceptable (Hosmer–Lemeshow test $P > 0.05$).

Discussion

Predictive models of live birth for patients with POR were developed and compared using data from the ESPART study, the largest randomized controlled trial conducted to date, in patients with a POR diagnosis incorporating the EBC (Humaidan et al., 2017a). These are the first predictive models specifically for patients with POR, as all others have been developed for the general population with infertility. The analysis demonstrates that only a small number of predictors (age, PNO and markers of ovarian reserve [AMH or AFC]) are needed to predict live birth with reasonable precision.

Table 4 – Published ranges and proposed thresholds (univariate analysis) for age, AMH level, AFC and number of oocytes retrieved in a previous ART treatment cycle.

Variable	Literature range	Literature references	Proposed threshold	95% CI
Age (years)	39–41	(Chuang et al., 2003; Ferraretti et al., 2011; Vaegter et al., 2017)	≤40	38–41
AMH (ng/ml)	0.4–1.5	(Celik et al., 2012; Ferraretti et al., 2011; La Marca and Sunkara, 2014a)	<0.5	0.3–1.1
AFC	3–8	(Ferraretti et al., 2011)	<7	4–8
PNO	2–6	(Cai et al., 2013)	<2	2–3

AFC, antral follicle count; AMH, anti-Müllerian hormone; ART, assisted reproductive technology; PNO, number of oocytes retrieved in a previous ART treatment cycle.

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013

Table 5 – Comparison of discrimination and determination (Nagelkerke's determination coefficient) between the general model (including all the variables issued from other models), the initial model (based on continuous significant predictors), simplified model (constituted by the dichotomized variables) and PROsPeR, unadjusted or adjusted for centre-effect.

Model	Selection	Centre adjustment	R^2[a]	AUC[b]	95% CI	Calibration[c]
General	r-hFSH (*n* = 477)	Unadjusted	0.078	0.726	0.662, 0.789	0.911
		Adjusted	0.203	0.834	0.791, 0.873	0.992
Initial	r-hFSH (*n* = 477)	Unadjusted	0.051	0.711	0.631, 0.762	0.791
		Adjusted	0.195	0.715	0.649, 0.781	0.681
Simplified	r-hFSH (*n* = 477)	Unadjusted	0.087	0.713	0.646, 0.780	0.531
		Adjusted	0.161	0.840	0.793, 0.887	0.782
PROsPeR[d]	r-hFSH (*n* = 477)	Unadjusted	0.067	0.709	0.639, 0.779	0.482
		Adjusted	0.173	0.835	0.786, 0.883	0.693
	r-hFSH/r-hLH (*n* = 462)	Unadjusted	0.037	0.564	0.463, 0.662	0.244
		Adjusted	0.148	0.655	0.609, 0.711	0.525
	ESPART full (*n* = 939)	Unadjusted	0.046	0.627	0.588, 0.677	0.362
		Adjusted	0.162	0.676	0.613, 0.723	0.463

[a] Denotes the Nagelkerke determination coefficient adapted to binary end-points.
[b] AUC for the receiver operator characteristic curve of c Statistics.
[c] Reported by the Hosmer–Lemeshow test *P*-value.
[d] The validation calculation was performed on patients who received r-hFSH, patients who received r-hFSH plus r-hLH and the total ESPART study population.
AUC, area under the curve; PROsPeR, Poor Responder Outcome Prediction; r-hFSH, recombinant human FSH; r-hLH, recombinant human LH.

Justification for model simplification

A pair of models was initially developed using continuous variables, then simplified by the use of dichotomized variables. The streamlined model, PROsPeR, provides a score of 0, 1 or 2 according to the number of predictive conditions met (**Table 6**). High discrimination was observed between the scores (OR 0.33 per score), which means that the expected live-birth rate is reduced by one-third with each increasing score. Moreover, when used within a single centre, PROsPeR has a high discrimination ($AUC_{ROC} \cong 83.5\%$) and has acceptable calibration.

The use of dichotomized data increases the ease of use by clinicians, but is generally not recommended by statisticians as there can be loss of information and power, as well as the possibility of reaching erroneous conclusions if arbitrary thresholds are selected (Royston et al., 2006). The use of dichotomized variables, however, can be justified as follows. First, non-linear effects were observed for all three predictors, with a notable change in the live-birth rate around these thresholds in the ESPART study data (Humaidan et al., 2017a). In addition, the chosen thresholds were confirmed by the Youden Index (maximization of sensitivity and specificity), which remained stable across random subsampled groups. Finally, a similar determination was observed for the continuous and dichotomized models, suggesting minimum loss of information.

The proposed thresholds have been validated in previous studies and used in other predictive models (Alviggi et al., 2009; Cai et al., 2013; Celik et al., 2012; Chuang et al., 2003; Ferraretti et al., 2011; La Marca and Sunkara, 2014a; Toftager et al., 2017; Vaegter et al., 2017). Age is a key predictor of live birth, and is known to have a quadratic non-linear effect, with a slow decline observed up to age 40 years, followed by a rapid acceleration of the decline. An association has been found between an AMH level less than 0.5 ng/ml and low ovarian reserve, and is included in the EBC (Ferraretti et al., 2011). PNO is also included in the EBC and other predictive models (Rongieres *et al.*, 2015; McLernon *et al.*, 2016), and a threshold of PNO less than 2 highlights the benefit of having multiple embryos available to transfer. Nonetheless, clinical assessment should incorporate the totality of an individual's medical status when applying their data to PROsPeR.

PROsPeR comparison with existing models

Despite a greater number of predictors used by previous predictive models, and the reduction of the PROsPeR approach to only three scores, the PROsPeR score reached higher performances, with a discrimination up to 83% when adjusted for centres. These results have at least three explanations. First, the model was developed specifically for patients with POR, resulting in a more homogeneous population. Second, the model was limited to a single treatment

Table 6 – Definition of each PROsPeR score.

Strict ESHRE Bologna criteria (ESPART inclusion criteria) (Humaidan et al., 2017a)	Patients should meet at least two of the following criteria: • Advanced maternal age (≥40– < 41 years) • Previous ART cycle with ≤3 oocytes retrieved with a conventional stimulation protocol • AMH 0.12–1.3 ng/ml		
PROsPeR Score	0	1	2
Patients experienced with ART treatment	<40 years and PNO ≥2	≥40 years or PNO <2	≥40 years and PNO <2
Patients inexperienced with ART treatment	<40 years and AMH >0.5 ng/ml	≥40 years or AMH ≤0.5 ng/ml	≥40 years and AMH ≤0.5 ng/ml

AMH, anti-Müllerian hormone; ART, assisted reproductive technology; ESHRE, European Society of Human Reproduction and Embryology; PNO, number of oocytes retrieved in a previous ART treatment cycle; PROsPeR, Poor Responder Outcome Prediction.

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013

(r-hFSH), which removed any variance caused by the treatment effect. Finally, through adequate calibration, PROsPeR eliminates the centre variability.

Although these principles have not previously been applied to ART predictive models, they are commonly used in other pathologies, including oncology, human immunodeficiency virus and acquired immune deficiency syndrome (Barretina et al., 2012; Revell et al., 2013). Development of population and treatment-specific models enable comparison of different treatments and thus identify the medication associated with the highest probability of success for an individual patient. Application of these principles to ART treatment would provide a first step towards improving individualized ovarian stimulation.

Clinical implication

The results reported herein demonstrate a highly significant variation of the intercept of the model but no significant difference of the PROsPeR score's main effect between centres. These results have two essential clinical implications. First, the intercept of the model is the mean live-birth rate over all the centres for the patients with a PROsPeR score of 0. The between-centre variation confirms the considerable between-centre heterogeneity of the performance (Arvis et al., 2012), which necessitates a calibration of the prediction when used in one centre. The particular formulation of PROsPeR allows a simple calibration (**Table 7**); second, no difference was found for the PROsPeR score effect between centres (OR 0.33, meaning a deterioration of one-third in the live-birth rate for each successive score); therefore, this effect can be considered valid.

Consequently, PROsPeR can be rigorously and easily applied to predict live birth for patients treated with r-hFSH at a particular centre. Conversely, its strict application to another drug, assuming a deterioration in live-birth rate of one-third for each increasing score in PROsPeR, is not appropriate, as shown by its application to r-hFSH plus r-hLH. This was, however, expected, as a difference was observed between the two medications when accounting for baseline conditions (Humaidan et al., 2017b). Therefore, generally speaking, a specific prediction score should be applied for each drug.

Moreover, the suggested PROsPeR classification of patients with POR into three groups has a more important implication. Owing to the demonstrated centre invariance, the PROsPeR score may also be considered as a useful tool to predict live birth for other medications. As such, it may represent a standard tool that could facilitate the comparison of efficacy between different medications.

Table 7 – Calibrating PROsPeR to a specific treatment centre and examples of its use.

Calibration

The live birth rate observed for patients with POR receiving r-hFSH for ovarian stimulation (LB_c) at a centre during the previous year, and the proportion of patients meeting the criteria for PROsPeR scores of 0,1 and 2 (p1, p2, and p3, respectively) are identified from the treatment center database.

The expected live birth rates (LB_1, LB_2, and LB_3) corresponding to the three PROsPeR scores can be calculated as follows:

- $LB_1 = LB_c/(p1+p2/3 + p3/9)$
- $LB_2 = LB_1/3$
- $LB_3 = LB_1/9$

For example, if LB_c =12%, and p1, p2 and p3 are 0.33, 0.53 and 0.14, respectively, then:

- $LB_1 = 12/(0.33 + [0.53/3] + [0.14/9]) = 23.0\%$
- $LB_2 = 7.7\%$
- $LB_3 = 2.6\%$

Case 1

36-year-old woman with a 3-year history of infertility, who had a previous ART treatment cycle in which three oocytes were retrieved and has an AMH of 0.6 ng/ml.

This woman is experienced with ART treatment and PROsPeR, including age and number of oocytes retrieved in a previous ART treatment cycle, would therefore be used. She is <40 and has a previous cycle with ≥2 oocytes retrieved. This gives a PROsPeR score of 0.

The predicted live birth rate is 23.0%.

Case 2

39-year-old woman with a 2-year history of infertility and a previous ART treatment cycle in which one oocyte was retrieved.

This woman is experienced with ART and PROsPeR, including age and number of oocytes retrieved in a previous ART treatment cycle, would therefore be used. She is <40 but has a previous cycle with <2 oocytes retrieved. This gives a PROsPeR score of 1.

The predicted live birth rate is 7.7%

Case 3

42-year-old woman with a 6-month history of infertility, who is inexperienced with ART and has an AMH of 0.3 ng/ml.

This woman is inexperienced with ART treatment and PROsPeR, including age and AMH, would therefore be used. She is ≥40 and has an AMH ≤0.5 ng/mL. This gives a PROsPeR score of 2.

The predicted live birth rate is 2.6%

Case 4

35-year-old woman with a 1-year history of infertility, who is inexperienced with ART treatment and has an AMH of 1.4 ng/ml.

This woman does not meet the criteria for poor ovarian response and PROsPeR would therefore not be suitable in this case.

AMH, anti-Müllerian hormone; ART, assisted reproductive technology; POR, poor ovarian response; PROsPeR, Poor Responder Outcome Prediction; r-hFSH, recombinant human FSH.

PROsPeR and the European Society of Human Reproduction and Embryology Bologna Criteria

The PROsPeR concept was developed using independent mathematical optimization; however, the three identified predictors are highly clinically relevant and broadly used, as previously described. These findings also validate the selection of predictors included in the EBC, which were developed by expert consensus. Furthermore, the models demonstrate that a number of the conditions included in the EBC are also predictive of live birth, the most important end-point for ART treatment. They also demonstrate the heterogeneity of the population with POR, as addressed in a number of other scientific publications (Alviggi et al., 2016; Busnelli et al., 2015; Ferraretti and Gianaroli, 2014; Humaidan et al., 2017a; La Marca et al., 2015; Papathanasiou, 2014; Venetis, 2014; Yang et al., 2015; Younis et al., 2015) .

The heterogeneity exists because the EBC aims to identify any patient at risk of POR, irrespective of specific clinical characteristics that would define degrees of POR (Ferraretti et al., 2011). This was discussed by the POSEIDON Group, who proposed further stratification of impaired ovarian response moving from POR to a 'low prognosis' concept (Alviggi et al., 2016; Humaidan et al., 2016). More specifically, it was suggested that oocyte quality, i.e. age-related aneuploidy rate, and hyposensitivity to standard doses of gonadotrophins. i.e. unexpected low response, should be integrated with markers of ovarian reserve and PNO when defining a patient's prognosis (Alviggi et al., 2016; Humaidan et al., 2016). This classification identifies four groups according to their clinical characteristics (**Supplementary**

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013

Table S1). The patients included in the ESPART study would be considered as the poorest responders (similar to Group 4), and PROsPeR has sub-divided this group further. This highlights the complexity of the overall 'POR group' and the complementarity of PROsPeR and the POSEIDON Group's proposals. However, as the score subdivides patients with POR according to the number of specific baseline criteria they meet, patients who have a PROsPeR score of 0, i.e. they met the EBC but were not classified as severe for any of them, could be considered to be 'borderline' POR, and this classification may, therefore, include patients with a more favourable prognosis for live birth.

Using PROsPeR in clinical practice

Further to demonstrating the independence of the PROsPeR effect per centre, PROsPeR can be easily and rigorously adapted for a specific centre if the live-birth rate observed for patients with POR receiving r-hFSH for ovarian stimulation at the centre during the previous year, and the proportion of patients meeting the criteria for PROsPeR scores of 0, 1 and 2, are known. The calculation and examples of use are shown in Table 7. Where AMH data are not available at a centre, an AFC threshold of less than 7 can be used in its place.

Limitations

The models described herein have a number of limitations. They were developed using data from the ESPART study in the population receiving r-hFSH, who were selected by inclusion and exclusion criteria that were based on, but stricter than, the EBC. This may limit the applicability of the model to only patients who might meet the ESPART study inclusion criteria. Furthermore, the selection criteria are the likely reason for the overall live birth rates of 11.7% observed in the ESPART study, as patients with the worst reproductive diagnosis were excluded. The sample size was modest and, as ESPART was a randomized controlled trial, there may be some bias when generalizing the data to routine clinical practice. In addition, AMH was assessed by a central laboratory in the ESPART study, whereas AFC analysis was conducted independently at each study centre. The AFC analysis is, therefore, likely to reflect inter-operator variability. Furthermore, the predictive models were developed using data for a single treatment cycle; therefore, predictions for cumulative live birth rates over multiple cycles cannot be provided. In addition, in spite of the observation of superior discrimination compared with previously published models, owing to the small sample size used to develop the model, the 95% CI for the OR for the linear approximation was large (OR 0.33 per score; 95% CI 0.20 to 0.55). The models need to be externally validated using a large database of patients treated in a real-world context. This work is currently under way and will demonstrate whether or not the model is applicable to a broader range of patients with POR, and whether the proposed reduction factor of three is accurate.

Conclusions

In conclusion, predictive models for live birth in patients with POR were developed and compared based on data from the ESPART study. By studying one ovarian stimulation treatment in particular, the effect of baseline patient characteristics on live birth can be optimally accounted for when considering three groups of patients with deteriorating predicted live birth rate. Moreover, this calculation can be easily and rigorously adapted to one centre. The accuracy of this approach was observed to be better than more complex models and, owing to the simple calculation required, may be useful in routine clinical practice. External validation and application of this concept to other medications are currently ongoing.

Acknowledgement

Medical writing support was provided by Alexander Jones, inScience Communications, and was funded by Merck KGaA, Darmstadt, Germany. This analysis was funded by Merck KGaA, Darmstadt, Germany.

Appendix: Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.rbmo.2018.03.013.

ARTICLE INFO

Article history:
Received 18 July 2017
Received in revised form 15 March 2018
Accepted 16 March 2018
Declarations: PL has received consulting fees from Merck KGaA, Darmstadt, Germany. WC and JS are employees of the EMD Serono Research and Development Institute, Billerica, MA, USA, a business of Merck KGaA, Darmstadt, Germany. TDH is an employee of Merck KGaA, Darmstadt, Germany. CA has received unrestricted research grants and consulting fees from Merck KGaA, Darmstadt, Germany, as well as honoraria for lectures from Merck and IBSA. PH has received unrestricted research grants from Ferring, MSD and Merck as well as honoraria for lectures from MSD, Merck and IBSA.

Keywords:
Assisted reproductive technology
ESPART study
Live birth
Poor ovarian responders
Poor ovarian response
Predictive model

REFERENCES

Alviggi, C., Humaidan, P., Howles, C.M., Tredway, D., Hillier, S.G., 2009. Biological versus chronological ovarian age: implications for assisted reproductive technology. Reprod. Biol. Endocrinol. 7, 101.

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013

Alviggi, C., Andersen, C.Y., Buehler, K., Conforti, A., De Placido, G., Esteves, S.C., Fischer, R., Galliano, D., Polyzos, N.P., Sunkara, S.K., Ubaldi, F.M., Humaidan, P., 2016. A new more detailed stratification of low responders to ovarian stimulation: from a poor ovarian response to a low prognosis concept. Fertil. Steril. 105, 1452–1453.

American Society for Reproductive Medicine, 2016. FAQs – the psychological component of infertility. https://connect.asrm.org/mhpg/education/infertilityimpact?ssopc=1. (Accessed 3 April 2018).

Arvis, P., Lehert, P., Guivarc'h-Leveque, A., 2012. Simple adaptations to the Templeton model for IVF outcome prediction make it current and clinically useful. Hum. Reprod. 27, 2971–2978.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M.F., Monahan, J.E., Morais, P., Meltzer, J., Korejwa, A., Jane-Valbuena, J., Mapa, F.A., Thibault, J., Bric-Furlong, E., Raman, P., Shipway, A., Engels, I.H., Cheng, J., Yu, G.K., Yu, J., Aspesi, P., Jr., de Silva, M., Jagtap, K., Jones, M.D., Wang, L., Hatton, C., Palescandolo, E., Gupta, S., Mahan, S., Sougnez, C., Onofrio, R.C., Liefeld, T., MacConaill, L., Winckler, W., Reich, M., Li, N., Mesirov, J.P., Gabriel, S.B., Getz, G., Ardlie, K., Chan, V., Myer, V.E., Weber, B.L., Porter, J., Warmuth, M., Finan, P., Harris, J.L., Meyerson, M., Golub, T.R., Morrissey, M.P., Sellers, W.R., Schlegel, R., Garraway, L.A., 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603–607.

Busnelli, A., Papaleo, E., Del Prato, D., La Vecchia, I., Iachini, E., Paffoni, A., Candiani, M., Somigliana, E., 2015. A retrospective evaluation of prognosis and cost-effectiveness of IVF in poor responders according to the Bologna criteria. Hum. Reprod. 30, 315–322.

Cai, Q., Wan, F., Huang, K., Zhang, H., 2013. Does the number of oocytes retrieved influence pregnancy after fresh embryo transfer? PLoS ONE 8, e56189.

Celik, H., Bildircin, D., Guven, D., Cetinkaya, M.B., Alper, T., Batuoglu, A.S., 2012. Random anti-Mullerian hormone predicts ovarian response in women with high baseline follicle-stimulating hormone levels : anti-Mullerian hormone in poor responders in assisted reproductive treatment. J. Assist. Reprod. Genet. 29, 797–802.

Chuang, C.C., Chen, C.D., Chao, K.H., Chen, S.U., Ho, H.N., Yang, Y.S., 2003. Age is a better predictor of pregnancy potential than basal follicle-stimulating hormone levels in women undergoing in vitro fertilization. Fertil. Steril. 79, 63–68.

Eijkemans, M.J., Kersten, F.A., Lintsen, A.M., Hunault, C.C., Bouwmans, C.A., Roijen, L.H., Habbema, J.D., Braat, D.D., 2017. Cost-effectiveness of 'immediate IVF' versus 'delayed IVF': a prospective study. Hum. Reprod. 32, 999–1008.

Ferraretti, A.P., Gianaroli, L., 2014. The Bologna criteria for the definition of poor ovarian responders: is there a need for revision? Hum. Reprod. 29, 1842–1845.

Ferraretti, A.P., La Marca, A., Fauser, B.C., Tarlatzis, B., Nargund, G., Gianaroli, L., 2011. ESHRE consensus on the definition of 'poor response' to ovarian stimulation for in vitro fertilization: the Bologna criteria. Hum. Reprod. 26, 1616–1624.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under the receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

Harrell, F.E., Jr., Lee, K.L., Mark, D.B., 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. 15, 361–387.

Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression, second ed. John Wiley & Sons, Inc., New York.

Humaidan, P., Alviggi, C., Fischer, R., Esteves, S.C., 2016. The novel POSEIDON stratification of 'Low prognosis patients in Assisted Reproductive Technology' and its proposed marker of successful outcome. F1000Res. 5, 2911.

Humaidan, P., Chin, W., Rogoff, D., D'Hooghe, T., Longobardi, S., Hubbard, J., Schertz, J., 2017a. Efficacy and safety of follitropin alfa/lutropin alfa in ART: a randomized controlled trial in poor ovarian responders. Hum. Reprod. 32, 544–555.

Humaidan, P., Chin, W., Rogoff, D., D'Hooghe, T., Longobardi, S., Hubbard, J., Schertz, J., 2017b. Efficacy and safety of follitropin alfa/lutropin alfa in ART: a randomized controlled trial in poor ovarian responders. Hum. Reprod. 32, 1537–1538.

La Marca, A., Sunkara, S.K., 2014a. Individualization of controlled ovarian stimulation in IVF using ovarian reserve markers: from theory to practice. Hum. Reprod. Update 20, 124–140.

La Marca, A., Sunkara, S.K., 2014b. Reply: the two sides of the individualization of controlled ovarian stimulation. Hum. Reprod. Update 20, 614–615.

La Marca, A., Grisendi, V., Giulini, S., Sighinolfi, G., Tirelli, A., Argento, C., Re, C., Tagliasacchi, D., Marsella, T., Sunkara, S.K., 2015. Live birth rates in the different combinations of the Bologna criteria poor ovarian responders: a validation study. J. Assist. Reprod. Genet. 32, 931–937.

Lehert, P., Kolibianakis, E.M., Venetis, C.A., Schertz, J., Saunders, H., Arriagada, P., Copt, S., Tarlatzis, B., 2014. Recombinant human follicle-stimulating hormone (r-HFSH) plus recombinant luteinizing hormone versus r-HFSH alone for ovarian stimulation during assisted reproductive technology: systematic review and meta-analysis. Reprod. Biol. Endocrinol. 12, 17.

Leushuis, E., van der Steeg, J.W., Steures, P., Bossuyt, P.M., Eijkemans, M.J., van der Veen, F., Mol, B.W., Hompes, P.G., 2009. Prediction models in reproductive medicine: a critical appraisal. Hum. Reprod. Update 15, 537–552.

McLernon, D.J., Steyerberg, E.W., Te Velde, E.R., Lee, A.J., Bhattacharya, S., 2016. 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 355, i5735.

Mochtar, M.H., Danhof, N.A., Ayeleke, R.O., Van der Veen, F., van Wely, M., 2017. Recombinant luteinizing hormone (rLH) and recombinant follicle stimulating hormone (rFSH) for ovarian stimulation in IVF/ICSI cycles. Cochrane Database Syst. Rev. (5), CD005070.

Nelson, S.M., Lawlor, D.A., 2011. 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. 8, e1000386.

Oudendijk, J.F., Yarde, F., Eijkemans, M.J., Broekmans, F.J., Broer, S.L., 2012. The poor responder in IVF: is the prognosis always poor?: a systematic review. Hum. Reprod. Update 18, 1–11.

Papathanasiou, A., 2014. Implementing the ESHRE 'poor responder' criteria in research studies: methodological implications. Hum. Reprod. 29, 1835–1838.

Pencina, M.J., D'Agostino, R.B., Vasan, R.S., 2010. Statistical methods for assessment of added usefulness of new biomarkers. Clin. Chem. Lab. Med. 48, 1703–1711.

Porcu, G., Lehert, P., Colella, C., Giorgetti, C., 2013. Predicting live birth chances for women with multiple consecutive failing IVF cycles: a simple and accurate prediction for routine medical practice. Reprod. Biol. Endocrinol. 11, 1.

Revell, A.D., Wang, D., Wood, R., Morrow, C., Tempelman, H., Hamers, R.L., Alvarez-Uria, G., Streinu-Cercel, A., Ene, L., Wensing, A.M., DeWolf, F., Nelson, M., Montaner, J.S., Lane, H.C., Larder, B.A., 2013. Computational models can predict response to HIV therapy without a genotype and may reduce treatment failure in different resource-limited settings. J. Antimicrob. Chemother. 68, 1406–1414.

Rongieres, C., Colella, C., Lehert, P., 2015. To what extent does Anti-Mullerian Hormone contribute to a better prediction of live birth after IVF? J. Assist. Reprod. Genet. 32, 37–43.

Royston, P., Altman, D.G., Sauerbrei, W., 2006. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat. Med. 25, 127–141.

Santi, D., Casarini, L., Alviggi, C., Simoni, M., 2017. Efficacy of Follicle-Stimulating Hormone (FSH) alone, FSH + luteinizing hormone, human menopausal gonadotropin or FSH + human chorionic

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013

gonadotropin on assisted reproductive technology outcomes in the 'Personalized' medicine era: a meta-analysis. Front. Endocrinol. (Lausanne) 8, 114.
Smith, A.D., Tilling, K., Lawlor, D.A., Nelson, S.M., 2015. External validation and calibration of IVFpredict: a national prospective cohort study of 130,960 in vitro fertilisation cycles. PLoS ONE 10, e0121357.
Steyerberg, E.W., Eijkemans, M.J., Habbema, J.D., 2001. Application of shrinkage techniques in logistic regression analysis: a case study. Stat. Neerl. 55, 76–88.
Teixeira, D.M., Martins, W.P., 2014. The two sides of the individualization of controlled ovarian stimulation. Hum. Reprod. Update 20, 614.
Templeton, A., Morris, J.K., Parslow, W., 1996. Factors that affect outcome of in-vitro fertilisation treatment. Lancet 348, 1402–1406.
Toftager, M., Bogstad, J., Lossl, K., Praetorius, L., Zedeler, A., Bryndorf, T., Nilas, L., Pinborg, A., 2017. Cumulative live birth rates after one ART cycle including all subsequent frozen-thaw cycles in 1050 women: secondary outcome of an RCT comparing GnRH-antagonist and GnRH-agonist protocols. Hum. Reprod. 32, 556–567.
van Loendersloot, L.L., van Wely, M., Repping, S., van der Veen, F., Bossuyt, P.M., 2011. Templeton prediction model underestimates IVF success in an external validation. Reprod. Biomed. Online 22, 597–602.
Vaegter, K.K., Lakic, T.G., Olovsson, M., Berglund, L., Brodin, T., Holte, J., 2017. Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. Fertil. Steril. 107, 641–648, e642.
Venetis, C.A., 2014. The Bologna criteria for poor ovarian response: the good, the bad and the way forward. Hum. Reprod. 29, 1839–1841.
Verhaak, C.M., Lintsen, A.M., Evers, A.W., Braat, D.D., 2010. Who is at risk of emotional problems and how do you know? Screening of women going for IVF treatment. Hum. Reprod. 25, 1234–1240.
Yang, S., Chen, X., Zhen, X., Wang, H., Ma, C., Li, R., Liu, P., Qiao, J., 2015. The prognosis of IVF in poor responders depending on the Bologna criteria: a large sample retrospective study from China. Biomed Res. Int. 2015, 296173.
Yang, Y., Sun, X., Cui, L., Sheng, Y., Tang, R., Wei, D., Qin, Y., Li, W., Chen, Z.J., 2016. Younger poor ovarian response women achieved better pregnancy results in the first three IVF cycles. Reprod. Biomed. Online 32, 532–537.
Younis, J.S., Ben-Ami, M., Ben-Shlomo, I., 2015. The Bologna criteria for poor ovarian response: a contemporary critical appraisal. J. Ovarian Res. 8, 76.

Please cite this article in press as: Philippe Lehert, Wai Chin, Joan Schertz, Thomas D'Hooghe, Carlo Alviggi, Peter Humaidan, Predicting live birth for poor ovarian responders: the PROsPeR concept, Reproductive BioMedicine Online (2018), doi: 10.1016/j.rbmo.2018.03.013