National Institute for Health and Care Excellence

Draft for consultation

Fertility problems: assessment and treatment

[J] Fertility prediction models and IVF access

NICE guideline NGXXX

Evidence reviews underpinning recommendations 1.9.4, 1.9.5 and 1.9.7 in the NICE guideline

September 2025

Draft for consultation

This evidence review was developed by NICE

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN:

Contents

Fertility prediction models and IVF access 6
Review question 6
Introduction 6
Summary of the protocol 6
Methods and process 7
Clinical prediction model evidence 7
Summary of included studies 9
Summary of the evidence 31
Economic evidence 35
Summary of included economic evidence 35
Economic model 42
The committee's discussion and interpretation of the evidence 45
Recommendations supported by this evidence review 47
References – included studies 48
Appendices 52
Appendix A Review protocols 52
Review protocol for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 52
Appendix B Literature search strategies 59
Literature search strategies for review question: What is the likelihood of spontaneous conception when tubal catheterisation/cannulation is used for the treatment of proximal tubal obstruction? 59
Appendix C Clinical prediction model evidence study selection 68
Study selection for: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 68
Appendix D Evidence tables 69
Evidence tables for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 69
Appendix E Forest plots 70
Forest plots for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management;

intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 70

Appendix F GRADE tables 78

GRADE tables for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 78

Appendix G Economic evidence study selection 98

Study selection for: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 98

Appendix H Economic evidence tables 99

Economic evidence tables for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 99

Appendix I Economic model 100

Economic model for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 100

Cost-utility analysis of IVF provision on the NHS 100

Appendix J Excluded studies 183

Excluded economic studies for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 221

Appendix K Research recommendations – full details 222

Research recommendations for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)? 222

Fertility prediction models and IVF access

Review question

What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using:

- expectant management,
- intrauterine insemination (IUI),
- IVF with or without intracytoplasmic sperm injection (ICSI)?

Introduction

People with health-related fertility problems often would like to know their chances of a live birth before starting IUI or IVF treatment. Clinical prediction models that estimate the expected reproductive outcomes after fertility treatment cycles can be useful in making an informed choice. Clinical prediction models are evolving continuously with new models being developed. Therefore, the aim of this review is to determine if clinical prediction models can predict the chances of live birth for people with health-related fertility problems based on the latest available evidence. Furthermore, using selected prediction models, a new economic model was developed to estimate the cost-effectiveness of IVF treatment cycles compared to expectant management.

Summary of the protocol

See Table 1 for a summary of the Population, Prognostic factors, Intervention, Comparison and Outcome characteristics of this review.

Table 1: Summary of the protocol

Population	People with a health-related fertility problem
Prognostic factors	Clinical prediction models that include (or control for) at least 2 of the following core set of factors: • Female age • Duration of subfertility • Cause of subfertility • Pregnancy history External validation of clinical prediction models, including: • Hunault model • McLernon model • Nelson and Lawler model • Templeton model • van Loendersloot model
Intervention	• Expectant management • Intrauterine insemination (IUI) • IVF with or without intracytoplasmic sperm injection (ICSI)
Comparison	N/A
Outcome	**Critical** • Live birth (as defined by study) Data will be extracted on model performance: • Discrimination:

	○ Concordance (C) statistic or area under the curve (AUC) with 95% confidence interval ○ D-statistic and standard error • Calibration: ○ number of observed (O) and expected (E) events ○ total O:E ratio **Important** None

IVF: in vitro fertilisation; N/A: not applicable

For further details see the review protocol in appendix A.

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A and the methods document (supplementary document 1).

Where there was discrimination and calibration data from multiple external validation studies for the same clinical prediction model, estimates of discrimination (ability of the model to distinguish between people who had a live birth and those who did not) and calibration (agreement between predicted outcomes and observed outcomes) were meta-analysed.

A modified GRADE approach was used for this review. Risk of bias and applicability (indirectness) were assessed with the Prediction model Risk Of Bias ASsessment Tool (PROBAST). Inconsistency was assessed using the prediction interval. The judgement of precision was based on the confidence intervals (CIs) of the C-statistic and O:E ratios. The minimally important differences of 0.75 for the C-statistic and a range of 0.8-1.2 for the observed:expected (O:E) ratio were used (based on definitions of 'good' discrimination and calibration in Debray et al. 2017). For judging imprecision of the model performance data in terms of the C-statistic, outcomes were downgraded once or twice based on the number of these thresholds that were crossed by the CI: C-statistic <0.6 poor discrimination; C-statistic 0.6-0.75 possibly helpful discrimination; C-statistic >0.75 clearly useful discrimination.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

Clinical prediction model evidence

Included studies

Twenty studies were included for this review: 4 studies developed and/or validated clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using expectant management (Collins 1995, Hunault 2005, Nguyen 2022, Snick 1997); 1 study examined the predictive performance of a clinical prediction model for intrauterine insemination (IUI) (Hansen 2016); and 15 studies developed and/or validated clinical prediction models for IVF (Arvis 2012, Balachandren 2020, Bhattacharya 2021, Devroe 2020, Dhillon 2016, Hamdine 2015, La Marca 2021, Leijdekkers 2018, McLernon 2016, Meijerink 2016, Nelson 2011, Ratna 2023, Rongieres 2015, Sarais 2016, Smith 2015).

The included studies are summarised in Table 2.

Many of the included studies report data for more than 1 model and include a range of development and validation (internal and external) design components.

Of the 4 studies examining the predictive performance of clinical prediction models for expectant management: 1 study described model development with internal validation

(Collins 1995); 1 study reported performance data for model development only, and was validated using data from Collins 1995 (Snick 1997). Two studies described validation of the previously developed Hunault (2004) model (Hunault 2005, Nguyen 2022). To note, the Hunault (2004) model was developed using data from 2 of the model development studies included in this review (Collins 1995, Snick 1997) and another study that could not be included in this review as relevant model performance data was not reported (Eimers 1994). The Hunault (2004) model development study could also not be included as model performance data was not reported in an extractable format, but external validation of this model is captured in this review. Hunault 2005 also describes external validation of an alternative Hunault (2004) model that includes the post-coital test (PCT) variable. Nguyen 2022 reported model performance data for the addition of a new predictor (anti-Müllerian hormone, AMH) to the Hunault (2004) model.

One study (Hansen 2016) reported relevant performance data for a clinical prediction model for IUI and this study described model development only with no internal or external validation.

Of the 15 studies examining the predictive performance of clinical prediction models for IVF, 4 studies described validation of the previously developed Templeton (1996) model (Arvis 2012, Nelson 2011, Rongieres 2015, Smith 2015). Two of these studies also reported data for centre-specific fitting of the Templeton model based on exactly the same variables (Arvis 2012, Rongieres 2015). The original model development study of Templeton (1996) did not meet inclusion criteria for this review as only goodness-of-fit was reported. Two studies (Arvis 2012, Nelson 2011) additionally reported model development data for adding new predictors to the Templeton model. Furthermore, 1 study (Rongieres 2015) described external validation for the Templeton-Arvis model (developed in Arvis 2012), and 1 study (Smith 2015) described external validation for the Nelson and Lawler model (developed in Nelson 2011). One study also described model development and internal validation data for adding new predictors and removing predictors from the Templeton model with the aim of developing a 'simplified' model (Rongieres 2015). One study (McLernon 2016) described model development with internal validation of a pre-treatment model and a ‘post-treatment’ model that included data from the first fresh embryo transfer to update predictions for subsequent cycles. Three studies described validation of the McLernon pre-treatment model (Bhattacharya 2021, Leijdekkers 2018, Ratna 2023) and 2 studies described external validation of the post-treatment McLernon model (Leijdekkers 2018, Ratna 2023). One study (Leijdekkers 2018) also described model development and internal validation for updated versions of the McLernon pre-treatment and post-treatment models with additional new predictors included. Two studies described validation of the previously developed van Loendersloot (2013) model (Devroe 2020, Sarais 2016), and reported data for refitting the van Loendersloot model based on the same variables but with re-estimation of predictor weights for the participating clinic. One of these studies (Devroe 2020) additionally reported model development and internal validation data for adding new predictors (IVF variables from current cycle) to the van Loendersloot model. To note, the original van Loendersloot (2013) model development study did not meet inclusion criteria for this review as it modelled ongoing pregnancy, but the external validation studies used live birth instead of ongoing pregnancy as the outcome of the van Loendersloot model. Two studies reported model development and validation using separate data (Dhillon 2016, Meijerink 2016). Two studies described model development with internal validation (Balachandren 2020, Hamdine 2015), and 1 study described model development and internal validation of 3 models (La Marca 2021).

See the literature search strategy in appendix B and study selection flow chart in appendix C.

Excluded studies

Studies not included in this review are listed, and reasons for their exclusion are provided in appendix J.

1 Summary of included studies

2 Summaries of the studies that were included in this review are presented in Table 2.

3 **Table 2: Summary of included studies.**

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
Arvis 2012 Cohort (prospective); 1 infertility clinic France Study dates: January 2000-June 2011	IVF N participants: NR N cycles: 12901 Female age in years, mean (SD): NR, median 33 (IQR 30-36) Duration of subfertility in years, mean (SD): 5.45 (2.6) Cause of subfertility: 48% male factor infertility; 16% tubal; 15% unexplained; 11% ovulatory; 10% endometriosis Primary infertility (%):77	TM1 (and TM2)-Validation of Templeton model (with centre-specific fitting): • Duration of subfertility (1 year; 4 years; 7 years; 13 years) • Female age (quadratic and cubic polynomial components of age) • Pregnancy history (previous LB with IVF; previous non-LB with IVF; previous LB not by IVF; previous non-LB not by IVF) • Previous IVF treatment outcome (number of previous unsuccessful IVF attempts) • Cause of subfertility (tubal factor) Model development (TM + additional variables): • Duration of subfertility 1 year; 4 years; 7 years; 13 years) • Female age (quadratic and cubic polynomial components of age) • Female BMI (>26 or <18) • Pregnancy history (previous LB with IVF; previous non-LB with IVF; previous LB not by IVF; previous non-LB not by IVF) • Previous IVF treatment outcome (number of previous	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Fresh only IVF/ICSI: ICSI Protocol outcomes not reported for calibration. Hosmer-Lemeshow test p-value level reported for TM1 ($p<0.001$) and indicated model under-estimation of LB. Calibration slope reported for TM2 and model development and identical for both (slope 0.96, 95% CI 0.88-1.51). Calibration plot for TM2 indicated good calibration

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
		unsuccessful IVF attempts) • Cause of subfertility (tubal factor) • Markers of ovarian reserve (FSH>10) • Smoking (smokes or smoked in the past) • Year (from 2011) Timepoint when model used: Pre-treatment		
Balachandren 2020 Cohort (retrospective); 1 tertiary referral hospital UK Study dates: January 2014-December 2016	IVF N participants: NR N cycles: 516 Female age in years, mean (SD): NR, median 35 (IQR 32-37) Duration of subfertility in years, mean (SD): NR Cause of subfertility: 39% male factor infertility; 39% unexplained; 9% tubal; 7% endometriosis; 5% ovulatory; 2% other Primary infertility (%): 78	• Female age (36-37; 38-39; 40-42) • Markers of ovarian reserve (FSH>12; AMH>8.5) Timepoint when model used: Pre-treatment	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Complete (both fresh and frozen) IVF/ICSI: IVF or ICSI Other candidate predictors considered but not included in final model: BMI, AFC category, duration of infertility, cause of infertility, and previous pregnancy No calibration data reported
Bhattacharya 2021 Cohort (retrospective); HFEA database UK Study dates: January-	IVF N participants: 9589 N cycles: NR Female age in years, mean (SD): 34.1 (4) Duration of subfertility in years, mean	McLM1 - Validation of McLernon model (pre-treatment): • Duration of subfertility (continuous) • Female age (continuous, as restricted cubic spline) • Pregnancy history (previous pregnancy in couple yes/no)	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Complete (both fresh and frozen) IVF/ICSI: IVF or ICSI Protocol outcomes not reported for calibration. For the model without

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
December 2017	(SD): NR, median 3 (IQR 2-6) Cause of subfertility: 42% unexplained; 34% male factor infertility; 10% tubal; 9% anovulation; 6% endometriosis Primary infertility (%):59.5	• Cause of subfertility (tubal; male factor; unexplained; anovulatory) • Male factor infertility (yes/no) • Number of cycles (cycle number) • Treatment type (IVF or ICSI) • Year of first egg retrieval (as restricted cubic spline) Timepoint when model used: Pre-treatment		recalibration, the calibration slope (0.78, 95% CI 0.70-0.87), calibration intercept (-0.58, 95% CI -0.63 to -0.52) and calibration plot indicated model over-estimation of LB No model performance data reported for the recalibrated model
Collins 1995 Cohort (prospective); 11 infertility clinics Canada Study dates: April 1984-March 1987	Expectant management N participants: 2198 N cycles: NR Female age in years, mean (SD): 29.5 (4.2) Duration of subfertility in years, mean (SD): 3.49 (2.25) Cause of subfertility: Some participants appear in more than 1 diagnosis group: 45% ovulatory; 27% male factor fertility problem (20% oligospermia and 7% azoospermia); 26% unexplained; 26% tubal; 14% endometriosis Primary infertility (%): 78	• Duration of subfertility (≤36 months) • Female age (≤30 years) • Pregnancy history (secondary infertility) • Cause of subfertility (tubal; endometriosis) • Male factor (oligospermia or azoospermia) Timepoint when model used: Pre-treatment	• Discrimination: C-statistic (95% CI)	Combined untreated couples with observations before treatment for treated couples Other candidate predictors considered but not included in final model: Male income; laparoscopy status; ovulatory defect; previous male treatment; female income; previous female treatment; coital frequency No calibration data reported

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
Devroe 2020 Cohort (retrospective); 1 infertility clinic (university hospital) Belgium Study dates: 2010-2018	IVF N participants: 591 N cycles: 1281 Female age in years, mean (SD): 34.4 (4.5) Duration of subfertility in years, mean (SD): 3.4 (2.2) Cause of subfertility: 54% male infertility [24% endometriosis; unclear if primary cause of infertility] Primary infertility (%):66	VLM (and VLM2)-Validation of van Loendersloot model (with centre-specific fitting): • Duration of subfertility (continuous; 5 years) • Female age (continuous) • Pregnancy history (previous delivery yes/no) • Previous IVF treatment outcome (number of previous failed IVF/ICSI cycles: 0; 1; 2; 3; 4; 5) • Markers of ovarian reserve (FSH continuous; FSH≤10) • Male factor infertility (yes/no) • Embryo quality (morphological score of all embryos day 3 in previous cycle) • Embryo grade (8-cell embryo on day 3 in previous cycle yes/no; morulae at day 3 in previous cycle yes/no) • Number of oocytes retrieved (number of embryos after oocyte retrieval in previous cycle; ≥10 embryos after oocyte retrieval in previous cycle yes/no) • Endometriosis (yes/no) Model development (VLM + additional variables): • Duration of subfertility (continuous; 5 years) • Female age (continuous) • Pregnancy history (previous delivery yes/no)	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Complete (both fresh and frozen) IVF/ICSI: IVF or ICSI Protocol outcomes not reported for calibration. For VLM, the calibration slope (0.64, 95% CI 0.47-0.81), intercept (0.20, 95% CI -0.14-0.53), and calibration plot indicated model under-estimation of LB. For VLM2 and model development, the calibration slope (VLM2: 1, 95% CI 0.79-1.21; model development: 1, 95% CI 0.83-1.17), intercept (VLM2: 0, 95% CI -0.23-0.23; model development: 0, 95% CI -0.18-0.18) and calibration plot indicated good calibration

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
		• Previous IVF treatment outcome (number of previous failed IVF/ICSI cycles: 0; 1; 2; 3; 4; 5) • Markers of ovarian reserve (FSH continuous; FSH≤10) • Male factor infertility (yes/no) • Embryo quality (morphological score of all embryos day 3 in previous cycle) • Embryo grade (≥1 8-cell embryo on day 3 in previous cycle; ≥1 morula on day 3 in previous cycle; ≥1 8-cell embryo on day 3 in current cycle; ≥1 morula on day 3 in current cycle) • Number of oocytes retrieved (number of embryos after oocyte retrieval in previous cycle; number with ≥10 embryos after oocyte retrieval in previous cycle; number of embryos after oocyte retrieval in current cycle; number with ≥10 embryos after oocyte retrieval in current cycle) • Endometriosis (yes/no) Timepoint when model used: After 1st cycle		
Dhillon 2016 Cohort (retrospective); CARE database UK & Ireland Study dates: 2008-2012	IVF N participants: 9915 N cycles: 9915 Female age in years, mean (SD): 34.6 (5.4) Duration of subfertility in	• Duration of subfertility (0-4 years; ≥5 years) • Female age (≤36 years; >36 years) • BMI (continuous) • Pregnancy history (previous live birth; previous miscarriage) • Cause of subfertility (male factor; tubal	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Fresh only IVF/ICSI: IVF or ICSI Protocol outcomes not reported for calibration. For the model

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
	years, mean (SD): 2 (2) Cause of subfertility: Some participants appear in more than 1 diagnosis group: 57% male factor infertility; 30% unexplained; 15% tubal; 11% anovulation; 30% other (e.g. endometriosis, fibroids) Primary infertility (%): 84	factor; anovulation; unexplained; other) • Markers of ovarian reserve (AFC continuous; AFC squared) • Male factor infertility (yes/no) • Ethnicity (white; Asian; black; Chinese; other; not stated; mixed) Timepoint when model used: Pre-treatment		without recalibration, the calibration slope (1.13, 95% CI 0.89-1.37), intercept (-0.17, 95% CI -0.25 to -0.08) and calibration plot indicated model over-estimation of LB
Hamdine 2015 Cohort (retrospective); 1 infertility clinic (university hospital) Netherlands Study dates: 2006-2011	IVF N participants: 487 N cycles: 1363 Female age in years, mean (SD): 34.6 (4.3) Duration of subfertility in years, mean (SD): 3.2 (2.1) Cause of subfertility: 43% male factor; 36% unexplained; 15% tubal; 2% endometriosis; 5% other Primary infertility (%):70	• Female age (continuous) • Pregnancy history (primary infertility; secondary infertility) • Previous IVF treatment outcome (number of previous ART treatments) • Markers of ovarian reserve (AMH continuous) Timepoint when model used: Pre-treatment	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Complete (both fresh and frozen) IVF/ICSI: IVF or ICSI Other candidate predictors considered but not included in final model: BMI, smoking status, duration of infertility, cause of subfertility, number of previous pregnancies and deliveries, type of treatment (IVF or ICSI) No calibration data reported Model predicting cumulative birth rate after 1-6 cycles over a 1-year period
Hansen 2016	IUI	• Duration of subfertility	• Discrimination: C-statistic (95% CI)	Secondary analysis of RCT (Diamond 2015)

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
Cohort (retrospective); 12 university hospital infertility clinics US Study dates: 2010-2013	N participants: 900 N cycles: 2572 Female age in years, mean (SD): 32.2 (4.3) Duration of subfertility in years, mean (SD): 2.9 (2.1) Cause of subfertility: 100% unexplained Primary infertility (%): 80	(continuous; in months) • Female age (continuous) • Pregnancy history (prior loss yes/no) • Ovarian stimulation agent (clomifene citrate; letrozole; gonadotropin) • Income (<$50,000; ≥$50,000) [Cause of subfertility not a variable in model but unexplained for all participants] Timepoint when model used: Pre-treatment		Other candidate predictors considered but not included in final model: BMI (female), BMI (male), waist circumference, waist-to-hip ratio, ethnicity, race, education, smoking, alcohol use, total motile sperm, prior conception, prior parity, prior infertility therapy, AMH, Emotion domain score of the FertiQol Protocol outcomes not reported for calibration. Hosmer-Lemeshow test p-value (p=0.99) indicated good calibration Time period/number of cycles over which outcome predicted: Up to 4 cycles
Hunault 2005 Cohort (prospective); 2 university hospitals Netherlands Study dates: January 1998-August 2002	Expectant management N participants: 302 N cycles: NR Female age in years, mean (SD): NR, median 32 (IQR 29-35) Duration of subfertility in years, mean (SD): NR,	HM1 - Validation of Hunault 2004 model (without PCT): • Duration of subfertility (1 year; 2 years; 3-4 years; 5-6 years; 7-8 years) • Female age (21-25 years; 26-31; 32-35; 36-37; 38-39; 40-41) • Pregnancy history (secondary infertility; primary infertility) • Sperm motility (≥60%; 40-59%; 20-39%; 0-19%)	• Discrimination: C-statistic (95% CI) • Calibration: number of observed and expected events and O:E ratio	Hunault 2004 used data from Collins (1995), Snick (1997) and Eimers (1994) to develop model Time period over which outcome predicted: 1 year

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
	median 2 (IQR 1.5-3) Cause of subfertility: NR (inclusion criteria unexplained, mild male factor, or cervical factor) Primary infertility (%): 63	• Referral status (secondary-care couple; tertiary care couple) HM2 - Validation of Hunault 2004 model (with PCT): • Duration of subfertility (1 year; 2 years; 3-4 years; 5-6 years; 7-8 years) • Female age (21-25 years; 26-31; 32-35; 36-37; 38-39; 40-41) • Pregnancy history (secondary infertility; primary infertility) • Sperm motility (≥60%; 40-59%; 20-39%; 0-19%) • Referral status (secondary-care couple; tertiary care couple) • Post-coital test (normal; abnormal) Timepoint when model used: Pre-treatment		
La Marca 2021 Cohort (retrospective); HFEA database UK Study dates: 1991-2011	IVF Model 1: N participants: 57699 N cycles: 57699 Model 2: N participants: 52960 N cycles: 52960 Model 3: N participants: 50870 N cycles: 50870 Female age in years, mean (SD): NR Duration of subfertility in years, mean (SD): NR	Model 1 (pre-treatment): • Duration of subfertility (<1 year; 1-3; 4-6; 7-9; 10-12; >12) • Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) • Pregnancy history (female primary infertility yes/no) • Cause of subfertility (unknown; low sperm count only; ovulatory only; tubal disease only; other) Model 2 (current cycle variable added to pre-treatment model): • Duration of subfertility (<1 year; 1-3; 4-6; 7-9; 10-12; >12)	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Fresh only IVF/ICSI: IVF or ICSI Data reported separately for development and validation sample (random-split of sample), extracted data for the internal validation sample Protocol outcomes not reported for calibration. Hosmer-Lemeshow test p-value for

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
	Cause of subfertility: NR Primary infertility (%): NR	• Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) • Pregnancy history (female primary infertility yes/no) • Cause of subfertility (unknown; low sperm count only; ovulatory only; tubal disease only; other) • Number of oocytes retrieved (1-4; 5-9; 10-14; 15-19; ≥20) Model 3 (current cycle variables added to pre-treatment model): • Duration of subfertility (<1 year; 1-3; 4-6; 7-9; 10-12; >12) • Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) • Pregnancy history (female primary infertility yes/no) • Cause of subfertility (unknown; low sperm count only; ovulatory only; tubal disease only; other) • Number of oocytes retrieved (1-4; 5-9; 10-14; 15-19; ≥20) • Number of embryos created (1-4; 5-9; 10-14; 15-19; ≥20) Timepoint when model used: Model 1: Pre-treatment Model 2: During 1st cycle (after oocyte retrieval) Model 3: During 1st cycle (after embryos created)		model 1 (p=0.2) and model 2 (p=0.8) indicated good calibration. For model 3, the p-value (p=0.01) indicated an imperfect fit
Leijdekkers 2018 Cohort (retrospective);	IVF N participants: 1511 N cycles: 2881	McLM1 - Validation of McLernon model (pre-treatment): • Duration of subfertility (continuous, range	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Complete (both fresh and frozen)

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
25-27 infertility clinics Netherlands Study dates: May 2011-May 2014	Female age in years, mean (SD): 33.5 (5) Duration of subfertility in years, mean (SD): NR. McLM1 & McLM2: median 2 (IQR 2-3) Model development 1: median 3 (IQR 2-4) Model development 2: median 4 (IQR 2-5) Cause of subfertility: 56% male factor; 35% unexplained; 11% tubal factor; 4% endometriosis Primary infertility (%): NR	of possible values 0-21) • Female age (continuous, range of possible values 18-50) • Pregnancy history (previous pregnancy in couple yes/no) • Cause of subfertility (tubal factor; anovulation; male factor; unexplained) • Male factor infertility (yes/no) • Number of cycles (cycle number) • Treatment type (IVF or ICSI) • Year of first egg retrieval (as restricted cubic spline) McLM2 - Validation of McLernon model (post-treatment): • Duration of subfertility (continuous, range of possible values 0-21) • Female age (continuous, range of possible values 18-50) • Pregnancy history (primary infertility of couple yes/no) • Cause of subfertility (tubal factor yes/no) • Embryo grade (1st fresh embryo transfer: no embryos transferred; single cleavage stage; single blastocyst stage; double cleavage stage; double blastocyst stage; triple cleavage stage; triple blastocyst stage) • Number of eggs collected in 1st cycle		IVF/ICSI: IVF or ICSI Data from OPTIMIST trial (van Tilborg 2017) used for external validation of McLM. External validation performed on cycles 1-3 (18 month follow-up) and data from cycles 4-6 excluded from analysis Protocol outcomes not reported for calibration. For external validation of the pre-treatment model (McLM1) without recalibration, the calibration slope (0.98, 95% CI 0.69-1.27), intercept (-0.23, 95% CI -0.36 to -0.10) and calibration plot indicated model over-estimation of LB. For external validation of the post-treatment model (McLM2), the calibration slope (0.97, 95% CI 0.77-1.19), intercept (-0.01, 95% CI -0.12-0.11) and calibration plot indicated good calibration Calibration data not reported for

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
		(range of possible values 1-28) • Fresh or frozen embryo transfer (embryos frozen in 1st cycle yes/no) • Number of cycles (cycle number) • Treatment type (IVF or ICSI in 1st cycle) • Year of first complete cycle (as restricted cubic spline) Model development pre-treatment (McLM1 + additional variables): • Duration of subfertility (continuous, range of possible values 0-21) • Female age (continuous, range of possible values 18-50) • BMI (weight, kg) • Pregnancy history (previous pregnancy in couple yes/no) • Cause of subfertility (tubal factor; anovulation; male factor; unexplained) • Markers of ovarian reserve (AMH, ng/l; AFC, 2-10mm) • Male factor infertility (yes/no) • Number of cycles (cycle number) • Treatment type (IVF or ICSI) • Year of first egg retrieval (as restricted cubic spline) Model development post-treatment (McLM2 + additional variables): • Duration of subfertility		model development

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
		(continuous, range of possible values 0-21) • Female age (continuous, range of possible values 18-50) • Pregnancy history (primary infertility of couple yes/no) • Cause of subfertility (tubal factor yes/no) • Markers of ovarian reserve (AMH, ng/l; AFC, 2-10mm) • Embryo grade (1st fresh embryo transfer: no embryos transferred; single cleavage stage; single blastocyst stage; double cleavage stage; double blastocyst stage; triple cleavage stage; triple blastocyst stage) • Number of eggs collected in 1st cycle (range of possible values 1-28) • Fresh or frozen embryo transfer (embryos frozen in 1st cycle yes/no) • Number of cycles (cycle number) • Treatment type (IVF or ICSI in 1st cycle) • Year of first complete cycle (as restricted cubic spline) Timepoint when model used: McLM1 & model development 1: Pre-treatment McLM2 & model development 2: After 1st cycle		
McLernon 2016	IVF	Model 1 (pre-treatment):	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Complete

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
Cohort (retrospective); HFEA database UK Study dates: January 1999-September 2008	N participants: 113873 N cycles: 184269 (complete cycles) Female age in years, mean (SD): 34.1 (5) Duration of subfertility in years, mean (SD): NR, median 4 (IQR 3-6) Cause of subfertility: 44% male factor; 29% unexplained; 23% tubal factor; 14% anovulatory; 7% endometriosis; 12% >1 type Primary infertility (%): NR	• Duration of subfertility (continuous) • Female age (continuous, as restricted cubic spline) • Pregnancy history (previous pregnancy in couple yes/no) • Cause of subfertility (tubal factor; anovulation; male factor; unexplained) • Male factor infertility (yes/no) • Number of cycles (cycle number) • Treatment type (IVF or ICSI) • Year of first egg retrieval (as restricted cubic spline) Model 2 (post-treatment): • Duration of subfertility (continuous) • Female age (continuous, as restricted cubic spline) • Pregnancy history (primary infertility of couple yes/no) • Cause of subfertility (tubal factor yes/no) • Embryo grade (1st fresh embryo transfer: no embryos transferred; single cleavage stage; single blastocyst stage; double cleavage stage; double blastocyst stage; triple cleavage stage; triple blastocyst stage) • Number of eggs collected in 1st cycle (as restricted cubic spline)		(both fresh and frozen) IVF/ICSI: IVF or ICSI No meaningful data reported for calibration (reported calibration slope but without 95% CI, intercept or calibration plot, this could not be interpreted) Cumulative probability estimated over up to 6 complete cycles Women who had no eggs collected in 1st cycle excluded from post-treatment model

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
		• Fresh or frozen embryo transfer (embryos frozen in 1st cycle yes/no) • Number of cycles (cycle number) • Treatment type (IVF or ICSI in 1st cycle) • Year of first complete cycle (as restricted cubic spline) Timepoint when model used: Model 1: Pre-treatment Model 2: After 1st cycle		
Meijerink 2016 Cohort (retrospective); 1 infertility clinic (university hospital) Netherlands Study dates: Development: September 2007-May 2014 Validation: August 2007-September 2015	IVF N participants: Development: 526 Validation: 289 N cycles: Development: 1006 Validation: 553 Female age in years, mean (SD): Development: 32.4 (4.4) Validation: 33.7 (4.5) Duration of subfertility in years, mean (SD): Development: 3.46 (2.1) Validation: NR Cause of subfertility: 100% male factor Primary infertility (%): Development: 82.5 Validation: NR	• Female age (age squared) • [Cause of subfertility not included as factor but inclusion criteria male factor] • Sperm motility (oocytes injected with motile; immotile; both motile and immotile) • Male factor infertility (suspected diagnosis before sperm retrieval: obstructive azoospermia; non-obstructive azoospermia) • Number of cycles (cycle number) • Male hormones level (LH continuous; testosterone continuous) Timepoint when model used: During 1st cycle (after oocyte retrieval)	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Complete (both fresh and frozen) IVF/ICSI: ICSI Inclusion criteria restricted to male factor fertility problems and ICSI Validation data collected from a different centre Other candidate predictors considered but not included in final model: Type of infertility (primary/secondary); duration of infertility (months); parity (n); average menstrual cycle length (days); uterine abnormalities (yes/no); antral follicle count before stimulation (number of follicles <11

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
				mm); alcohol use (self-reported; yes/no) for male and female; smoking status (self-reported; yes/no) for male and female; BMI at baseline (kg/m2) for male and female; male age (years); male inhibin B (ng/l); male FSH (IU/l); total testicular volume (cc); spermatozoa (fresh or frozen–thawed); number of oocytes retrieved Protocol outcomes not reported for calibration. Calibration slope (development: 1.07, 95% CI 0.89-1.25; validation: NR), intercept (development: 0.02, 95% CI -0.05-0.09; validation: NR), and p-value for Hosmer-Lemeshow test (development: p=0.79; validation: p=0.73) indicated good calibration
Nelson 2011 Cohort (retrospective); HFEA database UK	IVF N participants: NR N cycles: 144018 Female age in years, mean (SD): NR	TM1 - Validation of Templeton model: • Duration of subfertility (<1 year; 1-3 years; 4-6 years; 7-9 years; 9-12 years; >12 years)	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Fresh only IVF/ICSI: IVF or ICSI For model development

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
Study dates: January 2003-December 2007	Duration of subfertility in years, mean (SD): NR Cause of subfertility: NR Primary infertility (%): NR	• Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) • Pregnancy history (previous LB with IVF; previous non-LB with IVF; previous LB not by IVF; previous non-LB not by IVF) • Previous IVF treatment outcome (number of previous unsuccessful IVF attempts) • Cause of subfertility (tubal factor) Model development (TM + additional variables): • Duration of subfertility (<1 year; 1-3 years; 4-6 years; 7-9 years; 9-12 years; >12 years) • Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) • Pregnancy history (previous LB with IVF; previous non-LB with IVF; previous LB not by IVF; previous non-LB not by IVF) • Previous IVF treatment outcome (number of previous unsuccessful IVF attempts) • Cause of subfertility (unknown; tubal only; anovulatory only; endometriosis only; cervical only; male only; combination known causes) • Number of cycles (cycle number 1,2,3+) • Treatment type (IVF or ICSI) • Hormonal preparation		used 1000 bootstrap replications but estimates and SEs 'similar' and reported results without bootstrapping Calibration data not reported. Calibration plots indicated under-estimation of LB for TM1 and good calibration for model development

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
		(antioestrogen; gonadotropin; hormone replacement) • Source of egg (donor; participant) Timepoint when model used: TM: Pre-treatment Model development: During 1st cycle (after oocyte retrieval)		
Nguyen 2022 Cohort (retrospective); 1 private infertility clinic Australia Study dates: January-December 2012	Expectant management N participants: 325 N cycles: NR Female age in years, mean (SD): 31.7 (4.7) Duration of subfertility in years, mean (SD): NR, median 1.33 (IQR 1-2) Cause of subfertility: NR Primary infertility (%): 57.5	Validation of HM: • Duration of subfertility (continuous) • Female age (continuous) • Pregnancy history (secondary infertility; primary infertility) • Sperm motility (continuous) Model development (HM + additional variable): • Duration of subfertility (continuous) • Female age (continuous) • Pregnancy history (secondary infertility; primary infertility) • Sperm motility (continuous) • Markers of ovarian reserve (serum AMH ng/ml) Timepoint when model used: Pre-treatment	• Discrimination: C-statistic (95% CI) • Calibration: number of observed and expected events and O:E ratio	Calibration data only available for validation of HM (and not for model development of HM + AMH) Time period over which outcome predicted: 1 year
Ratna 2023 Cohort (retrospective); HFEA database UK	IVF N participants: 91035 N cycles: 144734 Female age in years, mean (SD): 35 (4)	McLM1 - Validation of McLernon model (pre-treatment): • Duration of subfertility (continuous, range of possible values 0-21) • Female age (continuous, range	• Discrimination: C-statistic (95% CI) • Calibration: O:E ratio	Definition of a cycle: Complete (both fresh and frozen) IVF/ICSI: IVF or ICSI For validation of post-treatment

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
Study dates: January 2010-December 2016	Duration of subfertility in years, mean (SD): NR, median 4 (IQR 2-6) Cause of subfertility: 39% male factor infertility; 31% unexplained; 15% tubal; 13% anovulation; 7% endometriosis; 12% >1 type Primary infertility (%):63	of possible values 18-50) • Pregnancy history (previous pregnancy in couple yes/no) • Cause of subfertility (tubal; male factor; unexplained; anovulatory) • Male factor infertility (yes/no) • Number of cycles (cycle number) • Treatment type (IVF or ICSI) • Year of first egg retrieval (as restricted cubic spline) McLM2 - Validation of McLernon model (post-treatment): • Duration of subfertility (continuous, range of possible values 0-21) • Female age (continuous, range of possible values 18-50) • Pregnancy history (previous pregnancy in couple yes/no) • Cause of subfertility (tubal factor yes/no) • Embryo grade (1st fresh embryo transfer: no embryos transferred; single cleavage stage; single blastocyst stage; double cleavage stage; double blastocyst stage; triple cleavage stage; triple blastocyst stage) • Number of eggs collected in 1st cycle (continuous, range of possible values 1-28)		model, data excluded from analysis for women with no oocytes collected in 1st treatment cycle (separate McLernon model for this group) Calibration slope (McLM1: 0.74, 95% CI 0.72-0.76; McLM2: 0.68, 95% CI 0.67-0.7) and calibration in the large (McLM1: 0.014; McLM2: -0.119) also reported

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
		• Fresh or frozen embryo transfer (embryos frozen in 1st cycle yes/no) • Number of cycles (cycle number) • Treatment type (IVF or ICSI in 1st cycle) • Year of first complete cycle (as restricted cubic spline) Timepoint when model used: McLM1: Pre-treatment McLM2: After 1st cycle		
Rongieres 2015 Cohort (retrospective); 1 infertility clinic France Study dates: Prior to December 2010	IVF N participants: NR N cycles: 715 Female age in years, mean (SD): NR, median 35.1 (IQR 31.1-38.7) Duration of subfertility in years, mean (SD): 1.93 (0.76) Cause of subfertility: 29% unexplained; 12% tubal; 7% IUI failure; 7% endometriosis; 10% age>40; 36% >1 type Primary infertility (%):60	TM1 (and TM2)- Validation of Templeton model (with centre-specific fitting): • Duration of subfertility (1 year; 4 years; 7 years; 13 years) • Female age (quadratic and cubic polynomial components of age) • Pregnancy history (previous LB with IVF; previous non-LB with IVF; previous LB not by IVF; previous non-LB not by IVF) • Previous IVF treatment outcome (number of previous unsuccessful IVF attempts) • Cause of subfertility (tubal factor) TMA- Validation of Templeton model: • Duration of subfertility 1 year; 4 years; 7 years; 13 years) • Female age (quadratic and cubic polynomial components of age)	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Fresh only IVF/ICSI: IVF or ICSI Protocol outcomes not reported for calibration. No calibration data reported for TM2 or model development. For TM1, the calibration plot and p-value for the Hosmer-Lemeshow test (p=0.008) indicated model under-estimation of LB. For TMA, the calibration slope (0.93, 95% CI -0.47-1.35) and calibration plot indicated good calibration

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
		• Female BMI (>26 or <18) • Pregnancy history (previous LB with IVF; previous non-LB with IVF; previous LB not by IVF; previous non-LB not by IVF) • Previous IVF treatment outcome (number of previous unsuccessful IVF attempts) • Cause of subfertility (tubal factor) • Markers of ovarian reserve (FSH>10) • Year (from 2011) Model development (TM +/- variables): • Pregnancy history (previous LB with IVF) • Markers of ovarian reserve (AMH continuous) • Year (from 2011) Timepoint when model used: Pre-treatment		
Sarais 2016 Cohort (retrospective); 1 infertility clinic Italy Study dates: January-December 2013	IVF N participants: 772 N cycles: 772 Female age in years, mean (SD): 36.9 (3.9) Duration of subfertility in years, mean (SD): 3.7 (1.3) Cause of subfertility: 30% male factor; 27% unexplained; 12% endometriosis;	VLM1 (and VLM2) - Validation of van Loendersloot model (with centre-specific fitting): • Duration of subfertility (≥5 years) • Female age (continuous) • Pregnancy history (previous ongoing pregnancy yes/no) • Previous IVF treatment outcome (number of previous failed IVF/ICSI cycles) • Markers of ovarian reserve (FSH≤10) • Male factor infertility (yes/no) • Embryo quality (morphological score	• Discrimination: C-statistic (95% CI)	Definition of a cycle: Complete (both fresh and frozen) IVF/ICSI: IVF or ICSI Protocol outcomes not reported for calibration. For VLM1, the calibration plot indicates under-estimation of LB. For VLM2, the calibration plot indicates good calibration

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
	6% tubal factor; 25% >1 type Primary infertility (%): 87	of all embryos day 3 in previous cycle) • Embryo grade (8-cell embryo on day 3 in previous cycle yes/no; morulae at day 3 in previous cycle yes/no) • Number of oocytes retrieved (≥10 embryos after oocyte retrieval in previous cycle yes/no) • Endometriosis (yes/no) Timepoint when model used: After 1st cycle		
Smith 2015 Cohort (retrospective); HFEA database UK Study dates: 2008-2010	IVF N participants: NR N cycles: 130960 Female age in years, mean (SD): NR Duration of subfertility in years, mean (SD): NR Cause of subfertility: 35% unexplained; 33% male factor; 12% tubal; 7% ovulatory; 4% endometriosis; 10% >1 type Primary infertility (%): 82	TM- Validation of Templeton model: • Duration of subfertility (<4 years; 4-6 years; 7-12 years; >12 years) • Female age (cubic polynomial components of age) • Pregnancy history (previous LB with IVF; previous non-LB with IVF; previous LB not by IVF; previous non-LB not by IVF) • Previous IVF treatment outcome (number of previous unsuccessful IVF attempts) • Cause of subfertility (tubal factor) NLM- Validation of Nelson Lawler model: • Duration of subfertility (<1 year; 1-3 years; 4-6 years; 7-9 years; 9-12 years; >12 years) • Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) • Pregnancy history (previous LB with IVF; previous non-LB with IVF;	• Discrimination: C-statistic (95% CI) • Calibration: number of observed and expected events and O:E ratio	Definition of a cycle: Fresh only IVF/ICSI: IVF or ICSI Paper also reports calibration slope (TM: 1.42, 95% CI 1.15-1.70; NLM: 0.93, 95% CI 0.84-1.03), intercept (TM: 0.08, 95% CI 0.04-0.12; NLM: 0.04, 95% CI 0.02-0.06), and p-value level for Hosmer-Lemeshow test (p<0.001)

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
		previous LB not by IVF; previous non-LB not by IVF) • Previous IVF treatment outcome (number of previous unsuccessful IVF attempts) • Cause of subfertility (tubal; anovulatory; endometriosis; cervical; male factor; combination known causes) • Number of cycles (cycle number 1,2,3+) • Treatment type (IVF or ICSI) • Hormonal preparation (antioestrogen; gonadotropin; hormone replacement) • Source of egg (donor; participant) Timepoint when model used: TM: Pre-treatment NLM: During 1st cycle (after oocyte retrieval)		
Snick 1997 Cohort (retrospective); 1 hospital Netherlands Study dates: January 1985-December 1993	Expectant management N participants: 726 N cycles: NR Female age in years, mean (SD): 29.1 (4.5) Duration of subfertility in years, mean (SD): 1.7 (1.2) Cause of subfertility: Some participants appear in more than 1 diagnosis group: 30%	Model 1: • Duration of subfertility (<24 months) • Cause of subfertility (tubal; ovulatory) • Post-coital test (abnormal) Model 2: • Duration of subfertility (<24 months) • Female age (<30 years) • Pregnancy history (secondary infertility) • Cause of subfertility (tubal; ovulatory) • Male factor (WHO semen abnormality)	• Discrimination: C-statistic (95% CI)	Combined untreated couples with observations before treatment for treated couples Paper refers to model 1 as selected model and model 2 as alternate model Model 2 performance data reported for model development dataset and validation dataset (validated using

Study	Intervention & population	Predictors/factors in model(s)	Outcomes	Comments
	unexplained; 30% male factor; 28% cervical; 26% ovulatory; 13% tubal; 3% endometriosis Primary infertility (%): 64	Timepoint when model used: Pre-treatment		data from Collins 1995) No calibration data reported

AFC: antral follicle count; AMH: anti-Müllerian hormone; ART: assisted reproductive technology; BMI: body mass index; C-statistic: Concordance statistic; CARE: Centres for assisted reproduction; CI: confidence interval; FertiQol: Fertility quality of life tool; FSH: follicle stimulating hormone; HFEA: Human fertilisation & embryology authority; HM: Hunault model; ICSI: intracytoplasmic sperm injection; IQR: interquartile range; IUI: intrauterine insemination; IVF: in vitro fertilisation; LB: live birth; LH: luteinizing hormone; McLM: McLernon model; ng/ml: nanograms per millilitre; NLM: Nelson Lawler model; NR: not reported; O:E: observed:expected; PCT: post-coital test; SD: standard deviation; SE: standard error; TM: Templeton model; TMA: Templeton-Arvis model; VLM: van Loendersloot model; WHO: World health organisation

See the full evidence tables in appendix D and the forest plots in appendix E.

Summary of the evidence

Expectant management

Hunault (2004) model

The Hunault (2004) clinical prediction model for expectant management included the following factors: duration of subfertility; female age; pregnancy history; sperm motility; referral status (secondary care or tertiary care). One version of the model also included post-coital test (PCT) as an additional factor.

Very low quality meta-analysed evidence from 2 external validation studies showed that the Hunault (2004) model (without the PCT factor) did not meet criteria for good discrimination (summary C-statistic estimate was 0.71, 95% confidence interval [CI] 0.007-1.0), and the meta-analysed O:E ratio (summary O:E estimate of 0.74, 95% CI 0.001-417.28) indicated imperfect calibration. There was very serious imprecision in both discrimination and calibration estimates. Very low quality evidence from a single external validation of the Hunault (2004) model with the PCT factor also did not show good discrimination performance (C-statistic 0.63, 95% CI 0.51-0.75), although calibration was good (O:E ratio 0.96, 95% CI not reported).

Low quality evidence from a model development study that added anti-Müllerian hormone (AMH) to the Hunault (2004) model, showed good discrimination (C-statistic 0.797, 95% CI 0.792-0.802) but calibration was not reported.

The model of Snick (1997) showed good discrimination, but this model was synthesised into the Hunault (2004) model. The Collins (1995) model did not meet criteria for good discrimination but again this was superseded by the Hunault (2004) model that used this data.

Intrauterine insemination (IUI)

Only 1 clinical prediction model was identified for IUI. This model was developed to predict cumulative live birth (over up to 4 cycles) prior to starting IUI with ovarian stimulation. The model included the following factors: duration of subfertility; female age; pregnancy history; ovarian stimulation agent; household income. Cause of subfertility was not included in the model as a factor but was controlled for as all participants had unexplained fertility problems.

Low quality evidence from this 1 model development study did not show good discrimination performance (C-statistic 0.65, 95% CI 0.61-0.69). The protocol outcomes for calibration (observed and expected events and observed:expected [O:E] ratio) were not reported, although the p-value for a statistical test for the goodness-of-fit (Hosmer-Lemeshow test) was reported and showed no statistically significant difference between the observed and expected events (p=0.99) indicating good calibration. However, no internal or external validation was described for this model.

In vitro fertilisation (IVF), with or without intracytoplasmic sperm injection (ICSI)

Templeton (1996) model

The Templeton (1996) model was developed to predict live birth prior to starting a cycle of IVF (with or without ICSI). The following factors were included: duration of subfertility; female age; pregnancy history; previous IVF treatment outcome; cause of subfertility.

Very low quality meta-analysed evidence from 4 external validation studies did not show good discrimination for the Templeton (1996) model (summary C-statistic estimate 0.62, 95% CI 0.61-0.63). It was not possible to meta-analyse the calibration estimates. The only study that reported the protocol outcome for calibration (observed and expected events and O:E ratio) showed poor calibration (O:E ratio 0.483, 95% CI 0.479-0.487). Three studies also reported significant p-values (p<0.01) on the Hosmer-Lemeshow test showing imperfect calibration. The calibration plots of all 4 studies indicated that the model underestimated live birth.

Very low quality meta-analysed evidence from 2 external validation studies reported data for centre-specific fitting of the Templeton model, but model performance still did not meet criteria for good discrimination (summary C-statistic estimate 0.69, 95% CI 0.61-0.76). Only 1 of these studies reported any information about calibration, with the calibration plot indicating good calibration.

Very low quality evidence from a model development study that produced a ‘simplified’ version of the Templeton (1996) model by adding and removing variables (predictors included in the model were pregnancy history, markers of ovarian reserve, and year) showed good discrimination (C-statistic 0.75, 95% CI 0.71-0.80). However, no calibration data was reported for this modified model, and no external validation.

Templeton-Arvis (2012) model

One study developed a new model (Templeton-Arvis [2012] model) for predicting live birth prior to starting a cycle of IVF with ICSI by extending the Templeton (1996) model with additional factors (markers of ovarian reserve; BMI; smoking; year).

Very low quality evidence from this model development study did not show good discrimination (C-statistic 0.71, 95% CI 0.68-0.74), and very limited data was reported for calibration.

Very low quality evidence from a single external validation study of the Templeton-Arvis (2012) model showed good discrimination (C-statistic 0.76, 95% CI 0.71-0.80). Calibration data was again limited but the calibration plot indicated good calibration.

Nelson and Lawler (2011) model

Another study developed a new model by adding predictors to the Templeton (1996) model (cycle number; treatment type [IVF or ICSI]; ovarian stimulation agent; source of egg). Very low quality evidence from this model development study did not show good discrimination (C-statistic 0.63, 95% CI 0.62-0.64). There was very limited calibration data available, although the calibration plot indicated good calibration.

Low quality evidence from an external validation study did not show good discrimination (C-statistic 0.628, 95% CI 0.625-0.631) for the Nelson and Lawler (2011) model. However, the model did show good calibration (O:E ratio 0.905, 95% CI 0.896-0.913).

McLernon (2016) model

The McLernon (2016) pre-treatment model included the following factors to predict the cumulative probability of live birth over up to 6 cycles of IVF (with or without ICSI): duration of subfertility; female age; pregnancy history; cause of subfertility; year; cycle number; treatment type (IVF or ICSI). The second McLernon (2016) model (post-treatment model) additionally included the following factors based on the first IVF cycle: oocytes retrieved; embryo quality/grade; fresh/frozen embryo.

Moderate quality evidence from the model development study showed discrimination performance fell below the threshold for good discrimination (C-statistic 0.73, 95% CI 0.72-0.74) for the pre-treatment model. Furthermore, no meaningful calibration data was reported (only the calibration slope with no 95% CI, intercept, or calibration plot). Low quality evidence from the same study also did not show good discrimination (C-statistic 0.72, 95% CI 0.71-0.73) and did not provide an interpretable calibration estimate for the post-treatment model.

Very low quality meta-analysed evidence from 3 external validation studies did not show good discrimination for the McLernon (2016) pre-treatment model (summary C-statistic estimate 0.63, 95% CI 0.54-0.72). It was not possible to meta-analyse the calibration estimates as not all studies reported calibration data and those that did used different measures. The only study that reported the protocol outcome for calibration (observed and expected events and O:E ratio) showed good calibration (O:E ratio 1.008, 95% CI 1.007-1.009). However, the calibration slope, intercept, and plots reported by the other 2 studies for the pre-treatment model (without recalibration) indicated model over-estimation of live birth.

Very low quality meta-analysed evidence from 2 external validation studies showed model performance that was below threshold for good discrimination for the McLernon (2016) post-treatment model (summary C-statistic estimate 0.73, 95% CI 0.43-0.91). Only 1 study reported the O:E ratio and this showed good calibration (O:E ratio 0.938, 95% CI 0.937-0.938). The calibration slope, intercept and plot reported by the other study also indicated good calibration for the post-treatment model.

Another study developed new pre-treatment and post-treatment models by adding predictors to the McLernon (2016) models. For the pre-treatment model, the new factors added were markers of ovarian reserve and BMI. Moderate quality evidence from this model development study did not show good discrimination (C-statistic 0.66, 95% CI 0.64-0.68) for this extended pre-treatment model, and no calibration data was reported. For the post-treatment model, the new factor added was markers of ovarian reserve and low quality evidence again did not show good discrimination (C-statistic 0.71, 95% CI 0.69-0.73) and no information about calibration for the modified post-treatment model was reported.

van Loendersloot (2013) model

The van Loendersloot (2013) model was developed to predict live birth before each IVF (with or without ICSI) cycle and included information from the previous cycle as factors in the model. The following factors were included: duration of subfertility; female age; pregnancy history; previous IVF treatment outcome; cause of subfertility; markers of ovarian reserve; embryos created (in previous cycle); embryo quality/grade (in previous cycle).

Very low quality meta-analysed evidence from 2 external validation studies did not show good discrimination for the van Loendersloot (2013) model (summary C-statistic estimate 0.64, 95% CI 0.64-0.64). Neither study reported the protocol outcomes for calibration, but the calibration slope and intercept reported by 1 study, and the plots reported by both studies indicated model under-estimation of live birth.

Very low quality meta-analysed evidence from these 2 external validation studies reported data for centre-specific fitting of the van Loendersloot model, and also did not show good discrimination (summary C-statistic estimate 0.66, 95% CI 0.39-0.86). Neither study reported the protocol outcomes for calibration, but the calibration slope and intercept reported by 1 study, and the plots reported by both studies indicated good calibration.

One of these studies developed a modified version of the van Loendersloot model that used current cycle, rather than previous cycle, variables (for number of embryos created and embryo quality/grade). Very low quality evidence from this study did not show good discrimination (C-statistic 0.71, 95% CI 0.68-0.75), although the calibration slope, intercept and plot indicated good calibration.

Model development and validation using separate data

One study developed a model to predict live birth prior to starting IVF (with or without ICSI) with information from a database for people undergoing treatment between 2008 and 2012, and the model was validated using database information from 2013. The following factors were included in the model: duration of subfertility; female age; pregnancy history; cause of subfertility; markers of ovarian reserve; BMI; ethnicity. High quality evidence from this study did not show good discrimination (C-statistic 0.62, 95% CI 0.60-0.64). The calibration slope, intercept and plot also indicated model over-estimation of live birth.

Another study developed a model to predict live birth for people undergoing ICSI using testicular extracted sperm and the model included current cycle variables (after oocyte retrieval). The model was developed using data from 1 centre and was validated using data from another centre. Very low quality evidence for the development set did not show good discrimination (C-statistic 0.62, 95% CI 0.58-0.68), although the calibration slope, intercept, plot and Hosmer-Lemeshow test p-value indicated good calibration. Very low quality evidence for the validation set also did not show good discrimination (C-statistic 0.67, 95% CI 0.62-0.72), although the calibration plot and Hosmer-Lemeshow test p-value indicated good calibration.

Model development with internal validation

One study developed a model to predict live birth prior to starting IVF (with or without ICSI) that included female age and markers of ovarian reserve. Very low quality evidence from this study did not show good discrimination (C-statistic 0.68, 95% CI 0.63-0.73). Furthermore, no calibration data was reported, and no external validation of this model.

Another study developed a model to predict cumulative live birth within 1 year of starting IVF (with or without ICSI). The model included the following factors: female age; pregnancy history; previous IVF treatment outcome; markers of ovarian reserve. Very low quality evidence from this study showed poor discrimination (C-statistic 0.59, 95% CI 0.56-0.63), and no calibration data was reported.

One study developed 3 models that predicted live birth using information from different stages of IVF (with or without ICSI) treatment. The first model predicted live birth prior to starting treatment and included the following factors: duration of subfertility; female age; pregnancy history; cause of subfertility. Very low quality evidence from this study did not show good discrimination (C-statistic 0.61, 95% CI 0.60-0.61), although the Hosmer-Lemeshow test p-value indicated good calibration. Model 2 added current cycle variables to the model after oocyte retrieval (number of oocytes retrieved), and very low quality evidence again did not show good discrimination (C-statistic 0.67, 95% CI 0.66-0.67) although the Hosmer-Lemeshow test p-value indicated good calibration for model 2. Model 3 added current cycle variables at the embryo creation stage (number of embryos obtained) and very low quality evidence did not show good discrimination (C-statistic 0.65, 95% CI 0.64-0.65), and the Hosmer-Lemeshow test p-value (p=0.01) indicated imperfect calibration for model 3.

See appendix F for modified GRADE tables.

Economic evidence

Included studies

A systematic review of the economic literature was conducted but no economic studies were identified which were applicable to this review question.

A total of 1,385 studies were identified in the health economic search for this review question. After duplicates were removed, 1,034 studies were sifted on title and abstract. Of these studies, all were excluded when sifting on title and abstract.

Excluded studies

Economic studies not included in this review are listed, and reasons for their exclusion are provided in appendix J.

Summary of included economic evidence

See Table 3 for the economic evidence profile for the model developed for the guideline

1 **Table 3: Economic evidence profile of a systematic review of economic evaluations of IVF access**

Study	Limitations	Applicability	Other comments	Incremental Costs Effect Cost effectiveness	Uncertainty
NICE guideline model 2025 **Cause of subfertility**: unknown	Potentially serious limitations[1,2, 3]	Directly applicable[4]	Effectiveness estimated form prediction models: **IVF:** OPIS pre-IVF tool **Spontaneous conception:** van Eekelen For results see Figure 22and Figure 23	**Age ≤ 37** IVF6 was <£20,000 per QALY **Age = 38** IVF4 was <£20,000 per QALY IVF6 was <£30,000 per QALY **Age = 39** IVF2 was <£20,000 per QALY IVF6 was <£30,000 per QALY **Age = 40** IVF3 was <£30,000 per QALY	Model results were sensitive to choice of model, assumptions made about depletion of susceptibles, assumptions with respect to health state utility and cost of IVF

Study	Limitations	Applicability	Other comments	Incremental Costs Effect Cost effectiveness	Uncertainty
NICE guideline model 2025 **Cause of subfertility**: tubal	Potentially serious limitations[1,2]	Directly applicable3	Effectiveness estimated form prediction models: **IVF:** OPIS pre-IVF tool **Spontaneous conception:** Zero assumed For results see Figure 24 and Figure 25	**Age ≤ 38** IVF6 was <£20,000 per QALY **Age = 39** IVF3 was <£20,000 per QALY IVF6 was <£30,000 per QALY **Age = 40** IVF1 was <£20,000 per QALY IVF4 was <£30,000 per QALY **Age = 41** IVF1 was <£30,000 per QALY	Model results were sensitive to choice of model, assumptions made about depletion of susceptibles, assumptions with respect to health state utility and cost of IVF
NICE guideline model 2025	Potentially serious limitations[1,2]	Directly applicable3	Effectiveness estimated form prediction models:	**Age ≤ 38**	Model results were sensitive to choice of model, assumptions made about depletion of susceptibles,

Study	Limitations	Applicability	Other comments	Incremental Costs Effect Cost effectiveness	Uncertainty
Cause of subfertility: Anovulatory			**IVF:** OPIS pre-IVF tool **Spontaneous conception:** Zero assumed For results see Figure 26 and Figure 27	IVF6 was <£20,000 per QALY **Age = 39** IVF5 was <£20,000 per QALY IVF6 was <£30,000 per QALY **Age = 40** IVF2 was <£20,000 per QALY IVF6 was <£30,000 per QALY **Age = 41** IVF3 was <£30,000 per QALY **Age = 42** IVF1 was <£30,000 per QALY	assumptions with respect to health state utility and cost of IVF

Study	Limitations	Applicability	Other comments	Incremental Costs Effect Cost effectiveness	Uncertainty
NICE guideline model 2025 **Cause of subfertility**: male factor	Potentially serious limitations[1,2]	Directly applicable[3]	Effectiveness estimated form prediction models: **IVF:** OPIS pre-IVF tool **Spontaneous conception:** van Eekelen For results see Figure 28 and Figure 29	**Age ≤ 37** IVF6 was <£20,000 per QALY **Age = 38** IVF4 was <£20,000 per QALY IVF6 was <£30,000 per QALY **Age = 39** IVF2 was <£20,000 per QALY IVF6 was <£30,000 per QALY **Age = 40** IVF3 was <£30,000 per QALY	Model results were sensitive to choice of model, assumptions made about depletion of susceptibles, assumptions with respect to health state utility and cost of IVF

Study	Limitations	Applicability	Other comments	Incremental Costs Effect Cost effectiveness	Uncertainty
NICE guideline model 2025 **Cause of subfertility**: mild endometriosis	Potentially serious limitations[1,2]	Directly applicable3	Effectiveness estimated form prediction models: **IVF:** IVF Predict **Spontaneous conception:** van Eekelen For results see Figure 30 and Figure 31	**Age ≤ 39** IVF6 was <£20,000 per QALY	Model results were sensitive to choice of model, assumptions made about depletion of susceptibles, assumptions with respect to health state utility and cost of IVF [5]
NICE guideline model 2025 **Cause of subfertility**: Severe endometriosis	Potentially serious limitations[1,2]	Directly applicable3	Effectiveness estimated form prediction models: **IVF:** IVF Predict **Spontaneous conception:** Zero assumed For results see Figure 32 and Figure 33	**Age ≤ 39** IVF6 was <£20,000 per QALY **Age 40-42** IVF4 was <£30,000 per QALY	Model results were sensitive to choice of model, assumptions made about depletion of susceptibles, assumptions with respect to health state utility and cost of IVF
NICE guideline model 2025	Potentially serious limitations[1,2]	Directly applicable3	Effectiveness estimated form prediction models: **IVF:**	**Age ≤ 34** IVF1 was <£20,000 per QALY	Model results were sensitive to choice of model, assumptions made about depletion of susceptibles, assumptions with respect to health state utility and cost of IVF

Study	Limitations	Applicability	Other comments	Incremental Costs Effect Cost effectiveness	Uncertainty
Cause of subfertility: Severe cervical			IVF Predict **Spontaneous conception:** Zero assumed For results see Figure 34 and Figure 35	IVF6 was <£30,000 per QALY **Age 35-39** IVF1 was <£30,000 per QALY	
NICE guideline model 2025 **Cause of subfertility**: combined	Potentially serious limitations[1,2]	Directly applicable[3]	Effectiveness estimated form prediction models: **IVF:** IVF Predict **Spontaneous conception:** van Eekelen For results see Figure 36 and Figure 37	**Age ≤ 34** IVF6 was <£20,000 per QALY **Age 35-39** IVF4 was <£20,000 per QALY IVF6 was <£30,000 per QALY	Model results were sensitive to choice of model, assumptions made about depletion of susceptibles, assumptions with respect to health state utility and cost of IVF [6]

[1] *No probabilistic sensitivity analysis was conducted.*
[2] *IVF and spontaneous conception prediction models were developed in different populations*
[3] *IVF prediction models are prone to selection bias, especially at higher order cycles (especially 4-6 cycles)*
[4] *This analysis was conducted specifically for this review question*
[5] *At age 35 only 4 cycles of IVF are cost-effective at £20,000 per QALY, with the additional 2 cycles cost-effective at £30,000 per QALY. At age 42, a single cycle is cost-effective at £30,000 per QALY. These results are an artefact of the age banding used in IVF Predict when compared with a spontaneous conception prediction model where cumulative live birth rated decline continuously with age*
[6] *At age 35-36 only a single cycle of IVF is cost-effective at £20,000 per QALY, with the additional 5 cycles cost-effective at £30,000 per QALY. These results are an artefact of the age banding used in IVF Predict when compared with a spontaneous conception prediction model where cumulative live birth rated decline continuously with age*

Economic model

An original cost-utility analysis was undertaken to compare the following strategies:

1. Expectant management (EM) for the remainder of the woman's reproductive life without IVF (no IVF)
2. One cycle of IVF, followed by EM for the remainder of the woman's reproductive life if 1 full cycle of IVF was unsuccessful (IVF1)
3. Up to 2 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 2 cycles of IVF were unsuccessful (IVF2)
4. Up to 3 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 3 cycles of IVF were unsuccessful (IVF3).
5. Up to 4 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 4 cycles of IVF were unsuccessful (IVF4).
6. Up to 5 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 5 cycles of IVF were unsuccessful (IVF5).
7. Up to 6 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 6 cycles of IVF were unsuccessful (IVF6).

The model was restricted to a maximum of 6 cycles of IVF, reflecting the upper limit of cycles in prediction models for IVF utilised in this analysis. The model is summarised below with full details available in appendix I.

A simple Markov approach was used to estimate cumulative live births from starting treatment to the end of a woman's reproductive life. In each monthly cycle women would either transition to a health state of "live birth" or remain in a "no birth" state as illustrated in Figure 1. The model accounted for the fact that woman who did not achieve live birth from IVF might do subsequently through spontaneous conception.

Prediction models were used to estimate the probability of live birth in each Markov cycle. In the base case analysis, the OPIS pre-IVF tool was used to estimate probabilities of live birth for IVF and live birth from EM was estimated using van Eekelen prediction model. The van Eekelen model was also used to estimate live births arising from spontaneous conception following the completion of IVF. Published literature was used to estimate ovarian hyperstimulation syndrome (OHSS) and multiple birth.

Figure 1: Markov schematic to assess fertility treatments across a woman's reproductive life cycle

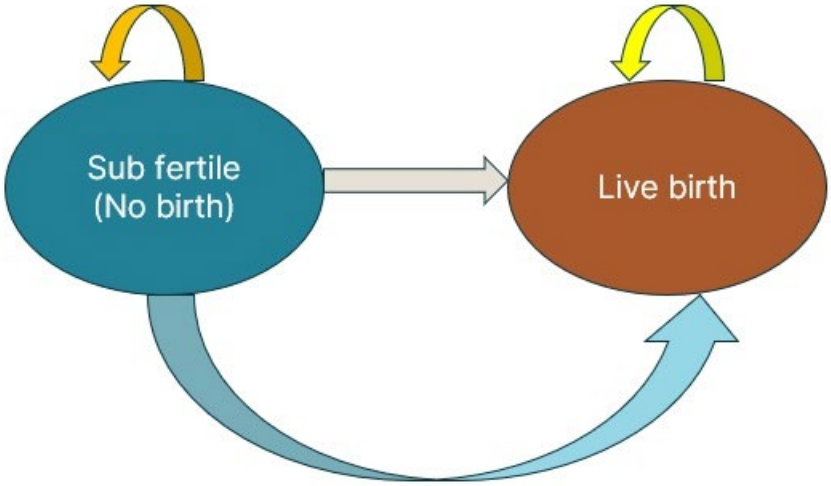
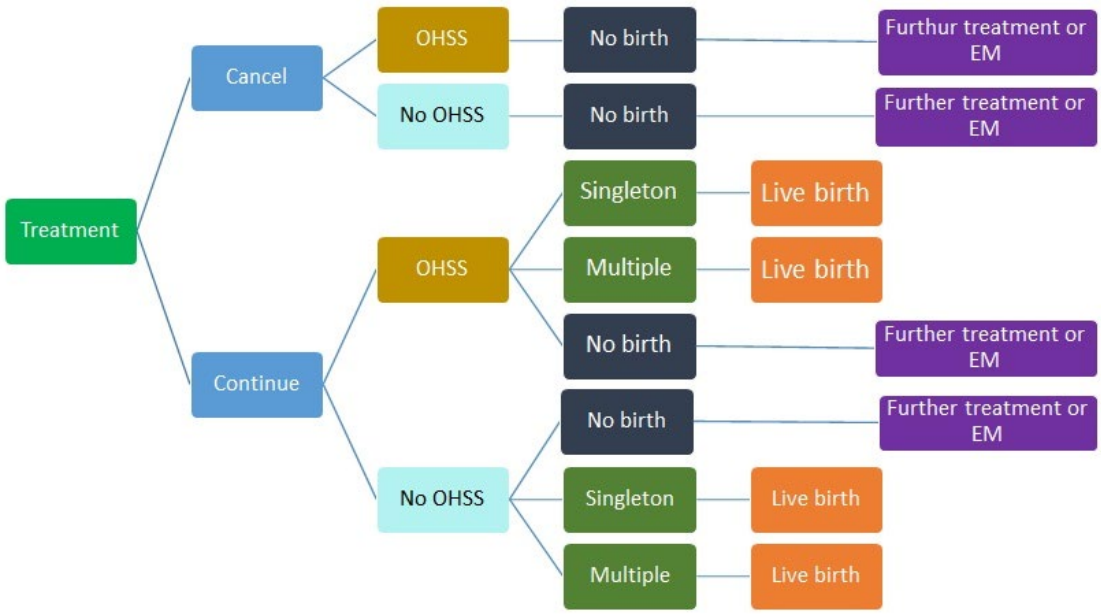


Figure 2 shows a schematic of the decision tree used to calculate the expected costs and outcomes associated with assisted reproduction.

Figure 2: Decision tree illustrating the outcomes of assisted reproduction

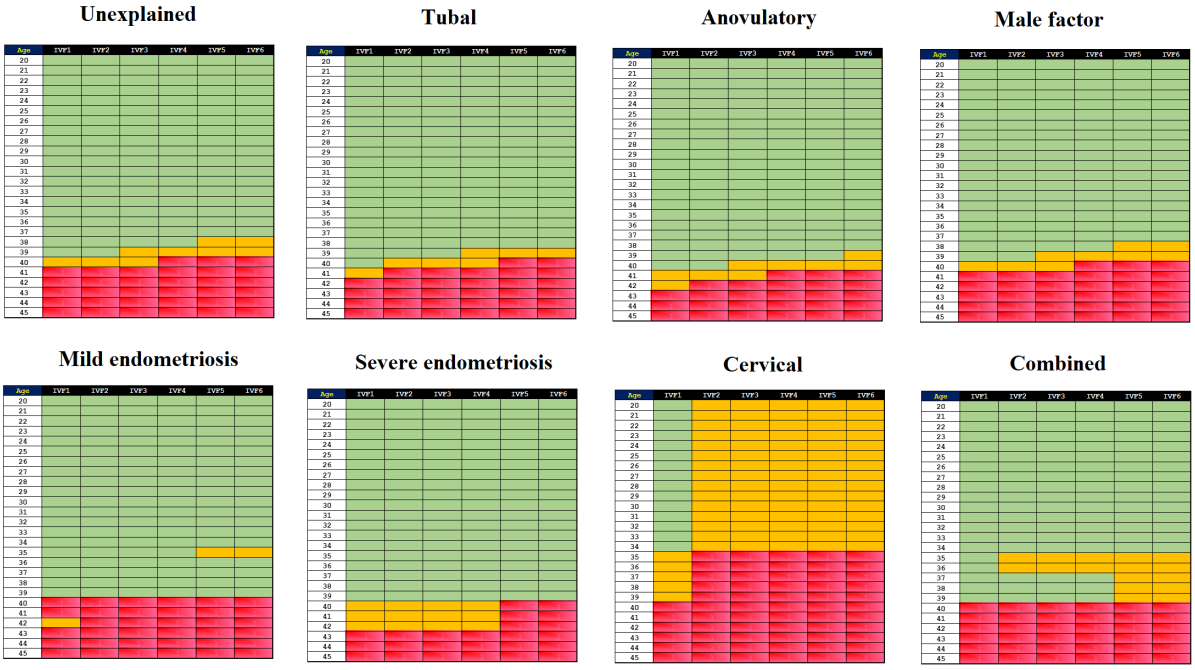


In addition to treatment costs the model also captured the "downstream" costs of multiple birth and OHSS associated with the different strategies. QALYs were estimated by assuming a health state utility gain from a live birth and then subtracting QALY losses from OHSS.

The results of the base case analysis are summarised graphically below in Figure 3. This economic analysis provides much stronger evidence for the cost-effectiveness of IVF compared with the previous NICE guideline (CG156), with lower ICERs attained in comparable analyses.

It suggested that 6 cycles of IVF were cost-effective for women aged 39 years and under. For women aged 40-41 years, the cost-effectiveness conclusions were more equivocal, but IVF was generally not found to be cost-effective for women aged 42 years and older. The model also provided some evidence that IVF with ICSI could be cost-effective for male factor cause despite the higher treatment cost.

Figure 3: Summary charts of cost-effectiveness of IVF cycles by age for the base case analysis using the OPIS pre-IVF (or IVF Predict) and van Eekelen prediction models



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

However, model cost-effectiveness conclusions were found to be sensitive to several assumptions or inputs over which there is uncertainty. In particular, the choice of prediction model, the assumed health state utility gain from live birth, the cost of IVF and accounting for depletion of susceptibles were all found to be important cost-effective drivers in some scenarios. Furthermore, there are legitimate concerns about potential selection bias in the prediction models as well as uncertainty relating to small samples especially at 4 cycles of IVF and above.

Therefore, while the committee believed that the cost-effectiveness evidence now supported the provision of 6 cycles to women aged 39 years and below, they recognised that the evidence was not sufficiently robust to make a strong recommendation for more than 3 cycles, although they did recommend that up to 6 cycles could be considered.

The evidence for the cost-effectiveness of IVF for women aged 40-41 was less convincing but there were a number of scenarios where at least 1 cycle was cost-effective and therefore they recommended that 1 IVF cycle be offered to these women, in line with existing NICE guidance. However, the committee believed that the models no longer supported the offer of IVF to women aged 42.

The committee's discussion and interpretation of the evidence

The outcomes that matter most

Live birth was prioritised as the critical outcome by the committee, as it is the best indicator of fertility, specified in the core outcome set for fertility research (Duffy 2020), and is the ultimate aim for people with fertility problems who are making fertility treatment decisions.

Model performance indicators that were selected as most informative were the C-statistic as this is the most commonly reported estimate of discrimination (ability of the model to distinguish between people who had a live birth and those who did not). The observed:expected (O:E) ratio was prioritised as the best measure of calibration (agreement between predicted outcomes and observed outcomes) as it could be meta-synthesised as a single measure (overcoming difficulties associated with the need to consider the calibration slope and intercept together), quantitatively captured the relationship presented in the calibration plot, and did not have the drawbacks associated with the Hosmer-Lemeshow test (for instance, no indication of the magnitude of the difference between expected and observed events, and the potential for being misleading with large sample sizes). However, the committee agreed that although they would prioritise the O:E ratio, where this was not reported, they would also consider other measures of calibration whilst bearing in mind the limitations of these measures.

The quality of the evidence

The quality of evidence was assessed using modified GRADE methodology. The evidence was predominately rated as low or very low quality, with only 1 high quality rating and a couple of moderate ratings across all model performance estimates. Common reasons for downgrading included risk of bias (assessed with the PROBAST checklist with risk of bias introduced by the analysis as the most common domain for rating at high risk), indirectness (serious or very serious concern about applicability of the evidence to the review question assessed with the PROBAST checklist; common concerns included strict inclusion/exclusion criteria, and model not intended to be used for first cycle), and imprecision due to 95% confidence intervals crossing decision making thresholds.

Prediction models

The committee discussed the model performance estimates for the clinical prediction models included in the review and noted that there was very little meta-analysable data (particularly for calibration), and no clinical prediction model met the pre-defined criteria for both good discrimination and good calibration. Based on the lack of any clear advantages for the use of clinical prediction models in counselling patients about their chances of a live birth, the committee agreed not to recommend the use of a clinical prediction model in routine clinical practice.

Cost effectiveness and resource use

No economic evidence was identified and therefore an original economic analysis was undertaken to support the committee in making cost-effectiveness recommendations. This analysis relied on prediction models which had a number of limitations, and the committee noted that these limitations would be propagated within the health economic model. Within the IVF prediction models it was noted that selection bias could potentially lead to over-estimates of treatment success especially at higher order of IVF cycles (especially 4-6 cycles). Furthermore, the datasets used to estimate IVF success had very small numbers having 4 or more cycles and therefore, in addition, there was considerable sampling uncertainty around the point estimates for predicted live birth rates for these higher order IVF cycles. The committee did note that predicted live birth rates for 1 and 3 cycles of IVF was

not inconsistent with RCT evidence but also that randomised evidence was lacking for more than 3 cycles of IVF.

The absolute treatment effect of IVF required a prediction model for expectant management (spontaneous conception leading to live birth) but the models were developed in different populations and the committee understood that there could be systematic differences in those populations and, in particular, that women referred for IVF may have a worse prognosis than women completing a fertility workup at a fertility clinic.

The committee also recognised that the QALY approach is not without its limitations and critics as an appropriate or sufficient measurement of benefit in the field of fertility treatment. Aside from the difficulties in the valuation of improvements in health-related quality of life arising from a live birth following subfertility, many also contend that there are other benefits, such as self-esteem and life satisfaction, which may in fact predominate.

Notwithstanding these limitations which may cause the cost-effectiveness to be either over or underestimated, the committee noted that the economic analysis undertaken for this guideline provided stronger evidence than the previous guideline (CG156) that IVF was a cost-effective treatment for the NHS to provide, especially for women aged 20-39 years old and using the preferred prediction models for IVF and expectant management. Indeed, the analysis suggested that 6 cycles could be considered cost-effective even when using a more restrictive cost-effectiveness threshold of £20,000 per QALY.

However, the committee acknowledged that increasing the number of cycles offered to women aged 20-39 years old from 3 to 6 cycles would have a significant resource impact to the NHS. Recognising that the model was particularly sensitive to a number of assumptions and inputs over which there was uncertainty, in addition to concerns about prediction model quality, patient selection bias and sampling uncertainty for higher order IVF cycles, the committee did not believe the evidence was sufficiently robust to make a strong recommendation to offer 6 cycles. Therefore, they made a strong recommendation that an initial 3 cycles of IVF be offered to women aged 20-39 years old but added that a further 3 cycles could be considered after discussing the likely probability of success and implications of further treatment with the patient. Based on their knowledge and experience, the committee did not think that many women would necessarily take up more IVF cycles even if they were offered.

The committee accepted that the model showed that for women aged 40-41 years, the cost-effectiveness of IVF was less clear cut. However, they did note that a more limited number of IVF cycles was often cost-effective especially if inherent uncertainty in model parameters was taken into account. The committee also believed that the evidence was not strong enough to remove a provision for IVF that currently exists and therefore they recommended that 1 cycle of IVF be offered to women in this age category.

However, the committee agreed that the health economic model no longer provided sufficiently strong evidence to support the provision of IVF to women aged 42 years old. Most analyses also did not indicate that IVF was cost-effective for women aged 42 years even when inputs were moved in a direction favouring IVF treatment. In women aged 42 years, no cycles of IVF were ever found to be cost-effective at an ICER of £20,000 per QALY in any of the base case analyses, and for most clinical scenarios included in the analysis, IVF was not cost-effective even when using the less stringent £30,000 per QALY decision threshold.

The committee noted that the apparent borderline cost-effectiveness of 1 cycle of IVF for mild endometriosis at age 42 was an artefact of having to use a prediction model that assumed the same IVF effectiveness across an age band of 40–42 years, which would lead to an underestimation of cost-effectiveness at 40 years and an overestimation at 42 years, because the reality is that IVF success is likely to diminish across that age band. The committee noted that the results for severe endometriosis were also based on an IVF prediction model that averaged IVF success across the 40–42 years age band and made a

simplifying assumption of a zero probability of spontaneous conception leading to live birth and which did not distinguish live birth rates by disease severity. For all these reasons, they believed the cost-effectiveness of IVF for women aged 42 years with severe endometriosis to be overestimated. The economic analysis did also suggest that 1 cycle of IVF might be cost-effective for anovulatory cause in women aged 42 years but the committee noted this was a very borderline finding using a more permissive £30,000 per QALY threshold. The committee also noted that IVF is not the first line treatment for anovulation and that anovulatory causes account for a small proportion of IVF. They also noted that women having IVF for anovulation are usually younger than 42 years of age.

The committee noted that the model also provided some evidence that IVF with ICSI could be cost-effective for male factor cause despite the higher treatment cost. They considered that this finding was consistent with existing recommendations on the indications for ICSI.

The committee accepted that their recommendations would likely have a significant resource impact to the NHS if implemented but only because recommendations in the previous NICE guideline were not fully implemented. The committee also believed that offering more cycles of IVF would not always lead to an increased acceptance of more IVF cycles, making the resource impact difficult to predict.

Recommendations supported by this evidence review

This evidence review supports recommendations 1.9.4, 1.9.5 and 1.9.7.

References – included studies

Clinical prediction model studies

Arvis 2012

Arvis P, Lehert P, Guivarc'h-Leveque A. Simple adaptations to the Templeton model for IVF outcome prediction make it current and clinically useful. Human Reproduction. 2012 27(10): 2971-8.

Balachandren 2020

Balachandren N, Salman M, Diu NL, Schwab S, Rajah K, Mavrelos D. Ovarian reserve as a predictor of cumulative live birth. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2020 252: 273-7.

Bhattacharya 2021

Bhattacharya S, Maheshwari A, Ratna MB, van Eekelen R, Mol BW, McLernon DJ. Prioritizing IVF treatment in the post-COVID 19 era: a predictive modelling study based on UK national data. Human Reproduction. 2021 36(3): 666-75.

Collins 1995

Collins JA, Burrows EA, Willan AR. The prognosis for live birth among untreated infertile couples. Fertility and Sterility. 1995 64(1): 22-8.

Devroe 2020

Devroe J, Peeraer K, Verbeke G, Spiessens C, Vriens J, Dancet E. 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(10): e037289.

Dhillon 2016

Dhillon RK, McLernon DJ, Smith PP, Fishel S, Dowell K, Deeks JJ, Bhattacharya S, Coomarasamy A. Predicting the chance of live birth for women undergoing IVF: a novel pretreatment counselling tool. Human Reproduction. 2016 31(1): 84-92.

Hamdine 2015

Hamdine O, Eijkemans MJ, Lentjes EG, Torrance HL, Macklon NS, Fauser BC, Broekmans FJ. Antimüllerian hormone: prediction of cumulative live birth in gonadotropin-releasing hormone antagonist treatment for in vitro fertilization. Fertility and Sterility. 2015 104(4): 891-8.

Hansen 2016

Hansen KR, He AL, Styer AK, Wild RA, Butts S, Engmann L, Diamond MP, Legro RS, Coutifaris C, Alvero R, Robinson RD. Predictors of pregnancy and live-birth in couples with unexplained infertility after ovarian stimulation–intrauterine insemination. Fertility and Sterility. 2016 105(6): 1575-83.

Hunault 2005

Hunault CC, Laven JS, van Rooij IA, Eijkemans MJ, te Velde ER, Habbema JD. Prospective validation of two models predicting pregnancy leading to live birth among untreated subfertile couples. Human Reproduction. 2005 20(6): 1636-41.

La Marca 2021

La Marca A, Capuzzo M, Donno V, Renzini MM, Del Giovane C, D'Amico R, Sunkara SK. The predicted probability of live birth in In Vitro Fertilization varies during important stages throughout the treatment: analysis of 114,882 first cycles. Journal of Gynecology Obstetrics and Human Reproduction. 2021 50(3): 101878.

Leijdekkers 2018

Leijdekkers JA, Eijkemans MJ, Van Tilborg TC, Oudshoorn SC, McLernon DJ, Bhattacharya S, Mol BW, Broekmans FJ, Torrance HL, OPTIMIST group. Predicting the cumulative chance of live birth over multiple complete cycles of in vitro fertilization: an external validation study. Human Reproduction. 2018 33(9): 1684-95.

McLernon 2016

McLernon DJ, Steyerberg EW, Te Velde ER, Lee AJ, Bhattacharya S. 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.

Meijerink 2016

Meijerink AM, Cissen M, Mochtar MH, Fleischer K, Thoonen I, De Melker AA, Meissner A, Repping S, Braat DD, van Wely M, Ramos L. Prediction model for live birth in ICSI using testicular extracted sperm. Human Reproduction. 2016 31(9): 1942-51.

Nelson 2011

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 Medicine. 2011 8(1): e1000386.

Nguyen 2022

Nguyen DK, O'Leary S, Gadalla MA, Roberts B, Alvino H, Tremellen KP, Mol BW. The predictive value of anti-Müllerian hormone for natural conception leading to live birth in subfertile couples. Reproductive BioMedicine Online. 2022 44(3): 557-64.

Ratna 2023

Ratna MB, Bhattacharya S, McLernon DJ. External validation of models for predicting cumulative live birth over multiple complete cycles of IVF treatment. Human Reproduction. 2023 38(10): 1998-2010.

Rongieres 2015

Rongieres C, Colella C, Lehert P. To what extent does anti-mullerian hormone contribute to a better prediction of live birth after IVF?. Journal of Assisted Reproduction and Genetics. 2015 32: 37-43.

Sarais 2016

Sarais V, Reschini M, Busnelli A, Biancardi R, Paffoni A, Somigliana E. Predicting the success of IVF: external validation of the van Loendersloot's model. Human Reproduction. 2016 31(6): 1245-52.

Smith 2015

Smith AD, Tilling K, Lawlor DA, Nelson SM. External validation and calibration of IVFpredict: a national prospective cohort study of 130,960 in vitro fertilisation cycles. PloS One. 2015 10(4): e0121357.

Snick 1997

Snick HK, Snick TS, Evers JL, Collins JA. The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. Human Reproduction. 1997 12(7): 1582-8.

Other

Alba 2017

Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, Guyatt G. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA. 2017 318(14): 1377-84.

Debray 2017

Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, Riley RD, Moons KG. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017 356.

Diamond 2015

Diamond MP, Legro RS, Coutifaris C, Alvero R, Robinson RD, Casson P, Christman GM, Ager J, Huang H, Hansen KR, Baker V. Letrozole, gonadotropin, or clomiphene for unexplained infertility. New England Journal of Medicine. 2015 373(13): 1230-40.

Eimers 1994

Eimers JM, te Velde ER, Gerritse R, Vogelzang ET, Looman CW, Habbema JD. The prediction of the chance to conceive in subfertile couples. Fertility and Sterility. 1994 61(1): 44-52.

Hunault 2004

Hunault CC, Habbema JD, Eijkemans MJ, Collins JA, Evers JL, Te Velde ER. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. Human Reproduction. 2004 19(9): 2019-26.

Templeton 1996

Templeton A, Morris JK, Parslow W. Factors that affect outcome of in-vitro fertilisation treatment. The Lancet. 1996 348(9039): 1402-6.

van Loendersloot 2013

Van Loendersloot LL, Van Wely M, Repping S, Bossuyt PM, Van Der Veen F. Individualized decision-making in IVF: calculating the chances of pregnancy. Human Reproduction. 2013 28(11): 2972-80.

van Tilborg 2017

Van Tilborg TC, Oudshoorn SC, Eijkemans MJ, Mochtar MH, Van Golde RJ, Hoek A, Kuchenbecker WK, Fleischer K, De Bruin JP, Groen H, Van Wely M. Individualized FSH dosing based on ovarian reserve testing in women starting IVF/ICSI: a multicentre trial and cost-effectiveness analysis. Human Reproduction. 2017 32(12): 2485-95.

Van Tilborg TC, Torrance HL, Oudshoorn SC, Eijkemans MJ, Koks CA, Verhoeve HR, Nap AW, Scheffer GJ, Manger AP, Schoot BC, Sluijmer AV. Individualized versus standard FSH

dosing in women starting IVF/ICSI: an RCT. Part 1: The predicted poor responder. Human Reproduction. 2017 32(12): 2496-505.

Appendices

Appendix A Review protocols

Review protocol for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

Table 4: Review protocol

ID	Field	Content
0.	PROSPERO registration number	CRD42023431351
1.	Review title	Predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems
2.	Review question	What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?
3.	Objective	To determine whether clinical prediction models can predict the chances of live birth for people with health-related fertility problems, to inform recommendations on criteria for access to treatments
4.	Searches	The following databases will be searched (with no date restrictions): Clinical searches • Cochrane Central Register of Controlled Trials (CENTRAL) • Cochrane Database of Systematic Reviews (CDSR) • Embase • MEDLINE ALL • Epistemonikos

ID	Field	Content
		Searches will be restricted by: • English language • Human studies The full search strategies for MEDLINE database will be published in the final review.
5.	Condition or domain being studied	Predictive models for the success of assisted reproduction techniques (ART) to inform recommendations on criteria for access to treatments
6.	Population	Inclusion: People with a health-related fertility problem. In this guideline, people with health-related fertility problems are those who have a known health-related impediment to fertility, or those who do not achieve a pregnancy: • after 12 months of regular unprotected sexual intercourse or • after 6 cycles of artificial insemination.
7.	Prognostic factors	Clinical prediction models that include (or control for) at least 2 of the following core set of factors: • Female age • Duration of subfertility • Cause of subfertility • Pregnancy history External validation of clinical prediction models, including: • Hunault model • McLernon model • Nelson and Lawler model • Templeton model • van Loendersloot model
8.	Comparator	N/A
9.	Types of study to be included	• Systematic reviews of clinical prediction models

ID	Field	Content
		• Primary studies that include the development and/or validation (internal or external) of clinical prediction models
10.	Other exclusion criteria	• Language limitations: non-English-language papers will be excluded • Conference abstracts, dissertations, and unpublished data will not be considered • Studies conducted in countries other than OECD high-income countries will be excluded (both development and validation samples should be in an OECD high income country)
11.	Context	This guidance will fully update the following NICE guideline: Fertility problems: assessment and treatment (last updated 2017; CG156)
12.	Primary outcomes (critical outcomes)	• Live birth (as defined by study) Data will be extracted on model performance: • Discrimination: o Concordance (C) statistic or area under the curve (AUC) with 95% confidence interval o D-statistic and standard error • Calibration: o number of observed (O) and expected (E) events o total O:E ratio Studies that do not include any of the above model performance data will be excluded
13.	Secondary outcomes (important outcomes)	N/A
14.	Data extraction (selection and coding)	All references identified by the searches and from other sources will be uploaded into EPPI and de-duplicated. Titles and abstracts of the retrieved citations will be screened to identify studies that potentially meet the inclusion criteria outlined in the review protocol. Dual sifting will be performed on at least 10% of records; 90% agreement is required. Disagreements will be resolved via discussion between the two reviewers, and consultation with senior staff if necessary. Full versions of the selected studies will be obtained for assessment. Studies that fail to meet the inclusion criteria once the full version has been checked will be excluded at this stage. Each study excluded after checking the full version will be listed, along with the reason for its exclusion.

ID	Field	Content	
		A standardised form will be used to extract data from studies. The following data will be extracted: study details (including reference, country, data source), number of participants, participant characteristics (data will be extracted on all key characteristics in line with the potential prognostic factors outlined above and the standard deviation of these values [to assess case mix]), inclusion and exclusion criteria, definition and measurement of the prognostic factors included in the model, definition of live birth, model development (including modelling method, method used to select variables for inclusion in the model, details on how non-linear relationships handled, and methods used for handling missing data), methods for model validation, model performance(calibration and discrimination), study classification according to TRIPOD checklist, setting, timing (including the timepoint when the clinical prediction model is used [e.g. pre-treatment, after first cycle] and the time period or number of cycles over which the outcome was predicted [the unit of analysis]), and source of funding. One reviewer will extract relevant data into a standardised form, and this will be quality assessed by a senior reviewer.	
15.	Risk of bias (quality) assessment	Quality assessment of individual studies will be performed using the following checklists: • ROBIS tool for systematic reviews • Prediction model Risk Of Bias ASsessment Tool (PROBAST) for primary studies that include the development and/or validation (internal or external) of clinical prediction models The quality assessment will be performed by one reviewer and this will be quality assessed by a senior reviewer.	
16.	Strategy for data synthesis	The predictive performance data will be quantitatively summarised, including summaries of discrimination (ability of the model to distinguish between people who conceived and those who did not) and calibration (agreement between predicted outcomes and observed outcomes). Where there is discrimination and calibration data from multiple external validation studies for the same clinical prediction model, estimates will be meta-analysed with a random-effects model (to allow for the presence of heterogeneity due to variability in design and population [case mix] of validation studies) using the metamisc package in R as outlined in Debray et al. 2019 (‘A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes’). Models will be included that are used to both predict outcome pre-treatment, and those that use information from the first cycle to update the model. However, these different timepoints will be differentiated in the analyses. ‘Good’ discrimination is defined as a C statistic >0.75 and ‘good’ calibration is defined as an O:E ratio between 0.8 and 1.2 (based on Debray et al. 2017).	
17.	Analysis of sub-groups	No planned subgroup analyses	
18.		☐	Intervention

<table>
<tr><th>ID</th><th>Field</th><th colspan="3">Content</th></tr>
<tr><td></td><td rowspan="6">Type and method of review</td><td colspan="2">☐</td><td>Diagnostic</td></tr>
<tr><td></td><td colspan="2">☒</td><td>Prognostic</td></tr>
<tr><td></td><td colspan="2">☐</td><td>Qualitative</td></tr>
<tr><td></td><td colspan="2">☐</td><td>Epidemiologic</td></tr>
<tr><td></td><td colspan="2">☐</td><td>Service Delivery</td></tr>
<tr><td></td><td colspan="2">☐</td><td>Other (please specify)</td></tr>
<tr><td>19.</td><td>Language</td><td colspan="3">English</td></tr>
<tr><td>20.</td><td>Country</td><td colspan="3">England</td></tr>
<tr><td>21.</td><td>Anticipated or actual start date</td><td colspan="3">November 2022</td></tr>
<tr><td>22.</td><td>Anticipated completion date</td><td colspan="3">November 2024</td></tr>
<tr><td>23.</td><td>Stage of review at time of this submission</td><td colspan="3"><table>
<tr><th>Review stage</th><th>Started</th><th>Completed</th></tr>
<tr><td>Preliminary searches</td><td>☐</td><td>☑</td></tr>
<tr><td>Piloting of the study selection process</td><td>☐</td><td>☑</td></tr>
<tr><td>Formal screening of search results against eligibility criteria</td><td>☐</td><td>☑</td></tr>
<tr><td>Data extraction</td><td>☐</td><td>☑</td></tr>
<tr><td>Risk of bias (quality) assessment</td><td>☐</td><td>☑</td></tr>
<tr><td>Data analysis</td><td>☐</td><td>☑</td></tr>
</table></td></tr>
<tr><td>24.</td><td>Named contact</td><td colspan="3">5a. Named contact
Guideline Development Team A

5b. Named contact e-mail
FertilityProblems@nice.org.uk</td></tr>
</table>

ID	Field	Content
		5c. Organisational affiliation of the review Guideline Development Team A, Centre for Guidelines, National Institute for Health and Care Excellence (NICE)
25.	Review team members	• Senior Technical Analyst • Technical Analyst
26.	Funding sources/sponsor	This systematic review is being completed by NICE
27.	Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.
28.	Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: https://www.nice.org.uk/guidance/indevelopment/gid-ng10263
29.	Other registration details	None
30.	Reference/URL for published protocol	https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=431351
31.	Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: • notifying registered stakeholders of publication • publicising the guideline through NICE's newsletter and alerts • issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.
32.	Keywords	Fertility problems, infertility, clinical prediction models, IUI, IVF, ICSI, expectant management

ID	Field	Content	
33.	Details of existing review of same topic by same authors	None	
34.	Current review status	☐	Ongoing
		☒	Completed but not published
		☐	Completed and published
		☐	Completed, published and being updated
		☐	Discontinued
35.	Additional information	None	
36.	Details of final publication	www.nice.org.uk	

1 *CDSR: Cochrane Database of Systematic Reviews; CENTRAL: Cochrane Central Register of Controlled Trials; GRADE: Grading of Recommendations Assessment,*
2 *Development and Evaluation; MEDLINE: Medical Literature Analysis and Retrieval System Online; N/A: not applicable; NICE: National Institute for Health and Care Excellence;*
3 *PROBAST: Prediction model Risk Of Bias ASsessment Tool; ROBIS: risk of bias in systematic reviews*

Appendix B Literature search strategies

Literature search strategies for review question: What is the likelihood of spontaneous conception when tubal catheterisation/cannulation is used for the treatment of proximal tubal obstruction?

Clinical search

Database: Medline – OVID interface

Date of last search: 11/10/2023

#	Searches
1	Fertilization In Vitro/ or exp Insemination, Artificial/ or Sperm Injections, Intracytoplasmic/
2	exp Infertility/th or Reproductive Techniques, Assisted/
3	Watchful Waiting/
4	Coitus/
5	(((artificial* or assist*) adj2 (conception* or reproduct*)) or ART).tw.
6	(IVF or (in vitro fertili* or invitro fertili*)).tw.
7	((fecund* or fertil* or hypofertil* or infertil* or steril* or subfecund* or subfertil*) adj2 (intervention* or therap* or treat*)).tw.
8	(((artificial* or assist* or intrauter* or intra uter*) adj2 inseminat*) or ICSI or IC SI or IUI).tw.
9	((intracytoplasm* or intra cytoplasm* or microinject* or micro inject* or transfer*) adj2 sperm*).tw.
10	((active* adj1 surveill*) or (clinical adj1 observ*) or (expect* adj2 (approach* or manag*)) or "no intervention*" or "no therap*" or "no treatment*" or (wait* adj1 see*) or (watch* adj2 wait*) or untreated).tw.
11	((coital or coitus or intercourse* or sex*) adj1 (frequen* or regular* or unprotect*)).tw.
12	or/1-11
13	Pregnancy/ or exp Pregnancy Outcome/ or exp Pregnancy Rate/
14	(livebirth* or (live adj1 birth*) or pregnan*).tw.
15	((baby or babies or child* or neonate* or newborn*) adj2 (birth* or born or conceiv* or deliver* or live* or living)).tw.
16	((birth* or conceiv* or conception* or gestation* or reproducti*) adj2 (achiev* or chance* or natural* or ongoing or outcome* or rate* or spontaneous* or success* or viab*)).tw.
17	or/13-16
18	12 and 17
19	letter/
20	editorial/
21	news/
22	exp historical article/
23	Anecdotes as topic/
24	comment/
25	case reports/
26	(letter or comment*).ti.
27	or/19-26
28	animals/ not humans/
29	exp Animals, Laboratory/
30	exp Animal Experimentation/
31	exp Models, Animal/
32	exp Rodentia/
33	(rat or rats or rodent* or mouse or mice).ti.
34	or/27-33
35	18 not 34
36	limit 35 to english language
37	Clinical Decision Rules/ or Models, Biological/ or Models, Theoretical/ or Models, Statistical/
38	Decision Support Systems, Clinical/ or Decision Support Techniques/ or Decision Making/
39	(decision* adj2 (clinical* or make or making or model* or support* or system* or technique*)).tw.
40	(Hunault* or Lawlor* or McLernon* or Nelson* or Templeton* or van Loendersloot*).tw.
41	(cumulati* adj3 (probab* or rate*)).tw.
42	(predict* or validat*).ti.
43	rule*.tw.
44	(predict* and (model* or outcome* or risk*)).ab.
45	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) and (decision* or identif* or model* or predict* or prognos*)).ti.
46	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) adj3 (decision* or identif* or model* or predict* or prognos*)).ab.
47	decision*.tw. and Logistic Models/
48	(prognostic and (characteristic* or criteria or factor* or finding* or history or model* or scor* or variable*)).tw.

#	Searches
49	ROC Curve/
50	(stratification or discrimination or discriminate or c statistic or area under the curve or AUC or calibration or index or indices or algorithm* or logistic regression or multivaria* or nomogram*).tw.
51	or/37-50
52	36 and 51

1 **Database: Embase – Ovid interface**

2 **Date of last search: 11/10/2023**

#	Searches
1	in vitro fertilization/ or exp artificial insemination/ or intracytoplasmic sperm injection/
2	infertility therapy/
3	watchful waiting/
4	sexual intercourse/
5	(((artificial* or assist*) adj2 (conception* or reproduct*)) or ART).tw.
6	(IVF or (in vitro fertili* or invitro fertili*)).tw.
7	((fecund* or fertil* or hypofertil* or infertil* or steril* or subfecund* or subfertil*) adj2 (intervention* or therap* or treat*)).tw.
8	(((artificial* or assist* or intrauter* or intra uter*) adj2 inseminat*) or ICSI or IC SI or IUI).tw.
9	((intracytoplasm* or intra cytoplasm* or microinject* or micro inject* or transfer*) adj2 sperm*).tw.
10	((active* adj1 surveill*) or (clinical adj1 observ*) or (expect* adj2 (approach* or manag*)) or "no intervention*" or "no therap*" or "no treatment*" or (without adj1 (intervention* or therap* or treatment*)) or (wait* adj1 see*) or (watch* adj2 wait*) or untreated).tw.
11	((coital or coitus or intercourse* or sex*) adj1 (frequen* or regular* or unprotect*)).tw.
12	or/1-11
13	live birth/
14	birth rate/ or pregnancy/ or pregnancy outcome/ or pregnancy rate/ or reproductive success/
15	(livebirth* or (live adj1 birth*) or pregnan*).tw.
16	((baby or babies or child* or neonate* or newborn*) adj2 (born or conceiv* or deliver* or live* or living)).tw.
17	((birth* or conceiv* or conception* or gestation* or reproducti*) adj2 (achiev* or chance* or natural* or ongoing or outcome* or rate* or spontaneous* or success* or viab*)).tw.
18	or/13-17
19	12 and 18
20	letter.pt. or letter/
21	note.pt.
22	editorial.pt.
23	case report/ or case study/
24	(letter or comment*).ti.
25	or/20-24
26	animal/ not human/
27	nonhuman/
28	exp Animal Experiment/
29	exp Experimental Animal/
30	animal model/
31	exp Rodent/
32	(rat or rats or rodent* or mouse or mice).ti.
33	or/25-32
34	19 not 33
35	limit 34 to english language
36	clinical decision rule/ or biological model/ or statistical model/ or theoretical model/
37	clinical decision making/ or clinical decision support system/ or decision making/ or decision support system/
38	(Hunault* or Lawlor* or McLernon* or Nelson* or Templeton* or van Loendersloot*).tw.
39	(decision* adj2 (clinical* or make or making or model* or support* or system* or technique*)).tw.
40	(cumulati* adj3 (probab* or rate*)).tw.
41	prediction/
42	(predict* or validat*).ti.
43	rule*.tw.
44	(predict* and (model* or outcome* or risk*)).ab.
45	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) and (decision* or identif* or model* or predict* or prognos*)).ti.
46	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) adj3 (decision* or identif* or model* or predict* or prognos*)).ab.
47	decision.tw. and exp statistical model/
48	(prognostic and (characteristic* or criteria or factor* or finding* or history or model* or scor* or variable*)).tw.
49	receiver operating characteristic/
50	(stratification or discrimination or discriminate or c statistic or area under the curve or AUC or calibration or index or indices or algorithm* or logistic regression or multivaria* or nomogram*).tw.

#	Searches
51	or/36-50
52	35 and 51
53	(conference abstract* or conference review or conference paper or conference proceeding).db,pt,su.
54	52 not 53

1

2 **Database: Cochrane Database of Systematic Reviews – Wiley interface**

3 **Date of last search: 11/10/2023**

ID	Search
#1	MeSH descriptor: [Fertilization in Vitro] this term only
#2	MeSH descriptor: [Insemination, Artificial] explode all trees
#3	MeSH descriptor: [Sperm Injections, Intracytoplasmic] this term only
#4	MeSH descriptor: [Infertility] explode all trees and with qualifier(s): [therapy - TH]
#5	MeSH descriptor: [Reproductive Techniques, Assisted] this term only
#6	MeSH descriptor: [Watchful Waiting] this term only
#7	MeSH descriptor: [Coitus] this term only
#8	(((artificial* or assist*) near/2 (conception* or reproduct*)) or ART):ti,ab
#9	(IVF or (("in vitro" or invitro) next fertili*)):ti,ab
#10	((fecund* or fertil* or hypofertil* or infertil* or steril* or subfecund* or subfertil*) near/2 (intervention* or therap* or treat*)):ti,ab
#11	(((artificial* or assist* or intrauter* or (intra next uter*) next/2 inseminat*)) or ICSI or IC SI or IUI):ti,ab
#12	((intracytoplasm* or (intra next cytoplasm*) or microinject* or (micro next inject*) or transfer*) near/2 sperm*):ti,ab
#13	((active* near/1 surveill*) or (clinical near/1 observ*) or (expect* near/2 (approach* or manag*)) or "no intervention" or "no interventions" or "no therapy" or "no therapies" or "no treatment" or "no treatments" or (wait* near/1 see*) or (watch* near/2 wait*) or untreated):ti,ab
#14	((coital or coitus or intercourse* or sex*) near/1 (frequen* or regular* or unprotect*)):ti,ab
#15	{or #1-#14}
#16	MeSH descriptor: [Pregnancy] this term only
#17	MeSH descriptor: [Pregnancy Outcome] explode all trees
#18	MeSH descriptor: [Pregnancy Rate] explode all trees
#19	(livebirth* or (live near1 birth*) or pregnan*):ti,ab
#20	((baby or babies or child* or neonate* or newborn*) near/2 (born or conceiv* or deliver* or live* or living)):ti,ab
#21	((birth* or conceiv* or conception* or gestation* or reproducti*) near/2 (achiev* or chance* or natural* or ongoing or outcome* or rate* or spontaneous* or success* or viab*)):ti,ab
#22	{or #16-#21}
#23	#15 and #22
#24	MeSH descriptor: [Clinical Decision Rules] this term only
#25	MeSH descriptor: [Models, Biological] this term only
#26	MeSH descriptor: [Models, Statistical] this term only
#27	MeSH descriptor: [Models, Theoretical] this term only
#28	MeSH descriptor: [Decision Support Systems, Clinical] this term only
#29	MeSH descriptor: [Decision Support Techniques] explode all trees
#30	MeSH descriptor: [Decision Making] this term only
#31	MeSH descriptor: [Logistic Models] this term only
#32	(decision* near/2 (clinical* or make or making or model* or support* or system* or technique*)):ti,ab
#33	(Hunault* or Lawlor* or McLernon* or Nelson* or Templeton* or van Loendersloot*):ti,ab
#34	(cumulati* near/3 (probab* or rate*)):ti,ab
#35	(predict* or validat*):ti
#36	rule*:ti,ab
#37	(predict* near/2 (model* or outcome* or risk*)):ab
#38	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) and (decision* or identif* or model* or predict* or prognos*)):ti
#39	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) near/2 (decision* or identif* or model* or predict* or prognos*)):ab
#40	(prognostic near/2 (characteristic* or criteria or factor* or finding* or history or model* or scor* or variable*)):ti,ab
#41	MeSH descriptor: [ROC Curve] this term only
#42	(stratification or discrimination or discriminate or "c statistic" or "area under the curve" or AUC or calibration or index or indices or algorithm* or "logistic regression" or multivaria* or nomogram*):ti,ab
#43	{or #24-#42}
#44	#23 and #43 in Cochrane Reviews, Cochrane Protocols

4

5 **Database Cochrane Central Register of Controlled Trials – Wiley interface**

1 **Date of last search: 11/10/2023**

ID	Search
#1	MeSH descriptor: [Fertilization in Vitro] this term only
#2	MeSH descriptor: [Insemination, Artificial] explode all trees
#3	MeSH descriptor: [Sperm Injections, Intracytoplasmic] this term only
#4	MeSH descriptor: [Infertility] explode all trees and with qualifier(s): [therapy - TH]
#5	MeSH descriptor: [Reproductive Techniques, Assisted] this term only
#6	MeSH descriptor: [Watchful Waiting] this term only
#7	MeSH descriptor: [Coitus] this term only
#8	(((artificial* or assist*) near/2 (conception* or reproduct*)) or ART):ti,ab
#9	(IVF or (("in vitro" or invitro) next fertili*)):ti,ab
#10	((fecund* or fertil* or hypofertil* or infertil* or steril* or subfecund* or subfertil*) near/2 (intervention* or therap* or treat*)):ti,ab
#11	(((artificial* or assist* or intrauter* or (intra next uter*) next/2 inseminat*)) or ICSI or IC SI or IUI):ti,ab
#12	((intracytoplasm* or (intra next cytoplasm*) or microinject* or (micro next inject*) or transfer*) near/2 sperm*):ti,ab
#13	((active* near/1 surveill*) or (clinical near/1 observ*) or (expect* near/2 (approach* or manag*)) or "no intervention" or "no interventions" or "no therapy" or "no therapies" or "no treatment" or "no treatments" or (wait* near/1 see*) or (watch* near/2 wait*) or untreated):ti,ab
#14	((coital or coitus or intercourse* or sex*) near/1 (frequen* or regular* or unprotect*)):ti,ab
#15	{or #1-#14}
#16	MeSH descriptor: [Pregnancy] this term only
#17	MeSH descriptor: [Pregnancy Outcome] explode all trees
#18	MeSH descriptor: [Pregnancy Rate] explode all trees
#19	(livebirth* or (live near1 birth*) or pregnan*):ti,ab
#20	((baby or babies or child* or neonate* or newborn*) near/2 (born or conceiv* or deliver* or live* or living)):ti,ab
#21	((birth* or conceiv* or conception* or gestation* or reproducti*) near/2 (achiev* or chance* or natural* or ongoing or outcome* or rate* or spontaneous* or success* or viab*)):ti,ab
#22	{or #16-#21}
#23	#15 and #22
#24	MeSH descriptor: [Clinical Decision Rules] this term only
#25	MeSH descriptor: [Models, Biological] this term only
#26	MeSH descriptor: [Models, Statistical] this term only
#27	MeSH descriptor: [Models, Theoretical] this term only
#28	MeSH descriptor: [Decision Support Systems, Clinical] this term only
#29	MeSH descriptor: [Decision Support Techniques] explode all trees
#30	MeSH descriptor: [Decision Making] this term only
#31	MeSH descriptor: [Logistic Models] this term only
#32	(decision* near/2 (clinical* or make or making or model* or support* or system* or technique*)):ti,ab
#33	(Hunault* or Lawlor* or McLernon* or Nelson* or Templeton* or van Loendersloot*):ti,ab
#34	(cumulati* near/3 (probab* or rate*)):ti,ab
#35	(predict* or validat*):ti
#36	rule*:ti,ab
#37	(predict* near/2 (model* or outcome* or risk*)):ab
#38	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) and (decision* or identif* or model* or predict* or prognos*)):ti
#39	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) near/2 (decision* or identif* or model* or predict* or prognos*)):ab
#40	(prognostic near/2 (characteristic* or criteria or factor* or finding* or history or model* or scor* or variable*)):ti,ab
#41	MeSH descriptor: [ROC Curve] this term only
#42	(stratification or discrimination or discriminate or "c statistic" or "area under the curve" or AUC or calibration or index or indices or algorithm* or "logistic regression" or multivaria* or nomogram*):ti,ab
#43	{or #24-#42}
#44	#23 and #43
#45	"conference":pt or (clinicaltrials or trialsearch):so
#46	#44 not #45 in Trials

2

3 **Database: Epistemonikos**

4 **Date of last search: 11/10/2023**

ID	Search
1	(((artificial* or assist*) and (conception* or reproduct*)) or ART or IVF or ("in vitro fertilisation" or "in vitro fertilization" or "invitro fertilisation" or "invitro fertiliation"))

ID	Search
2	(((artificial* or assist* or intrauter* or uter*) and inseminat*) or ICSI or "IC SI" or IUI or ((intracytoplasm* or cytoplasm* or microinject* or inject* or transfer*) and sperm*))
3	((fecund* or fertil* or hypofertil* or infertil* or steril* or subfecund* or subfertil*) and (intervention* or therap* or treat*))
4	((active* and surveill*) or (clinical and observ*) or (expect* and (approach* or manag*)) or "no intervention" or "no therapy" or "no treatment" or (wait* and see*) or (watch* and wait*))
5	(((coitus or coital) and (frequen* or regular* or unprotect*)) or ((sex* or intercourse*) and (frequen* or regular* or unprotect*)))
6	1 or 2 or 3 or 4 or 5
7	((birth* or conceiv* or conception* or gestation* or pregnan* or reproducti*) and (achiev* or chance* or live or living or natural* or ongoing or outcome* or rate* or spontaneous* or success* or viab*))
8	6 and 7
9	(((predict* or cumulat* or decision* or likelihood or prognos* or validat*) and (characteristic* or criteria or factor* or finding* or history or model* or outcome* or rate* or risk* or rule* or scor* or variable*)) or stratification or discrimination or discriminate or "c statistic " or "c-statistic " or "area under the curve " or AUC or calibration or index or indices or algorithm* or "logistic regression" or multivaria* or nomogram*)
10	8 and 9

1

2 **Economic literature search strategies:**

3 **Database: Medline – Ovid platform**

4 **Date last searched: 19/07/2023**

5

#	Searches
1	Fertilization In Vitro/ or exp Insemination, Artificial/ or Sperm Injections, Intracytoplasmic/
2	exp Infertility/th or Reproductive Techniques, Assisted/
3	Watchful Waiting/
4	Coitus/
5	(((artificial* or assist*) adj2 (conception* or reproduct*)) or ART).tw.
6	(IVF or (in vitro fertili* or invitro fertili*)).tw.
7	((fecund* or fertil* or hypofertil* or infertil* or steril* or subfecund* or subfertil*) adj2 (intervention* or therap* or treat*)).tw.
8	(((artificial* or assist* or intrauter* or intra uter*) adj2 inseminat*) or ICSI or IC SI or IUI).tw.
9	((intracytoplasm* or intra cytoplasm* or microinject* or micro inject* or transfer*) adj2 sperm*).tw.
10	((active* adj1 surveill*) or (clinical adj1 observ*) or (expect* adj2 (approach* or manag*)) or "no intervention*" or "no therap*" or "no treatment*" or (without adj1 (intervention* or therap* or treatment*)) or (wait* adj1 see*) or (watch* adj2 wait*) or untreated).tw.
11	((coital or coitus or intercourse* or sex*) adj1 (frequen* or regular* or unprotect*)).tw.
12	or/1-11
13	Pregnancy/ or exp Pregnancy Outcome/ or exp Pregnancy Rate/
14	(livebirth* or (live adj1 birth*) or pregnan*).tw.
15	((baby or babies or child* or neonate* or newborn*) adj2 (born or conceiv* or deliver* or live* or living)).tw.
16	((birth* or conceiv* or conception* or gestation* or reproducti*) adj2 (achiev* or chance* or natural* or ongoing or outcome* or rate* or spontaneous* or success* or viab*)).tw.
17	or/13-16
18	12 and 17
19	letter/
20	editorial/
21	news/
22	exp historical article/
23	Anecdotes as topic/
24	comment/
25	case reports/
26	(letter or comment*).ti.
27	or/19-26
28	animals/ not humans/
29	exp Animals, Laboratory/
30	exp Animal Experimentation/
31	exp Models, Animal/
32	exp Rodentia/
33	(rat or rats or rodent* or mouse or mice).ti.

#	Searches
34	or/27-33
35	18 not 34
36	limit 35 to english language
37	Clinical Decision Rules/ or Models, Biological/ or Models, Theoretical/ or Models, Statistical/
38	Decision Support Systems, Clinical/ or Decision Support Techniques/ or Decision Making/
39	(decision* adj2 (clinical* or make or making or model* or support* or system* or technique*)).tw.
40	(Hunault* or Lawlor* or McLernon* or Nelson* or Templeton* or van Loendersloot*).tw.
41	(cumulati* adj3 (probab* or rate*)).tw.
42	(predict* or validat*).ti.
43	rule*.tw.
44	(predict* and (model* or outcome* or risk*)).ab.
45	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) and (decision* or identif* or model* or predict* or prognos*)).ti.
46	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) adj3 (decision* or identif* or model* or predict* or prognos*)).ab.
47	decision*.tw. and Logistic Models/
48	(prognostic and (characteristic* or criteria or factor* or finding* or history or model* or scor* or variable*)).tw.
49	ROC Curve/
50	(stratification or discrimination or discriminate or c statistic or area under the curve or AUC or calibration or index or indices or algorithm* or logistic regression or multivaria* or nomogram*).tw.
51	or/37-50
52	36 and 51
53	Economics/
54	Value of life/
55	exp "Costs and Cost Analysis"/
56	exp Economics, Hospital/
57	exp Economics, Medical/
58	Economics, Nursing/
59	Economics, Pharmaceutical/
60	exp "Fees and Charges"/
61	exp Budgets/
62	budget*.ti,ab.
63	cost*.ti.
64	(economic* or pharmaco?economic*).ti.
65	(price* or pricing*).ti,ab.
66	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
67	(financ* or fee or fees).ti,ab.
68	(value adj2 (money or monetary)).ti,ab.
69	or/53-68
70	52 and 69

1

2 **Database: Embase**

3 **Date last searched: 19/07/2023**

#	Searches
1	in vitro fertilization/ or exp artificial insemination/ or intracytoplasmic sperm injection/
2	infertility therapy/
3	watchful waiting/
4	sexual intercourse/
5	(((artificial* or assist*) adj2 (conception* or reproduct*)) or ART).tw.
6	(IVF or (in vitro fertili* or invitro fertili*)).tw.
7	((fecund* or fertil* or hypofertil* or infertil* or steril* or subfecund* or subfertil*) adj2 (intervention* or therap* or treat*)).tw.
8	(((artificial* or assist* or intrauter* or intra uter*) adj2 inseminat*) or ICSI or IC SI or IUI).tw.
9	((intracytoplasm* or intra cytoplasm* or microinject* or micro inject* or transfer*) adj2 sperm*).tw.

#	Searches
10	((active* adj1 surveill*) or (clinical adj1 observ*) or (expect* adj2 (approach* or manag*)) or "no intervention*" or "no therap*" or "no treatment*" or (without adj1 (intervention* or therap* or treatment*)) or (wait* adj1 see*) or (watch* adj2 wait*) or untreated).tw.
11	((coital or coitus or intercourse* or sex*) adj1 (frequen* or regular* or unprotect*)).tw.
12	or/1-11
13	live birth/
14	birth rate/ or pregnancy/ or pregnancy outcome/ or pregnancy rate/ or reproductive success/
15	(livebirth* or (live adj1 birth*) or pregnan*).tw.
16	((baby or babies or child* or newborn*) adj2 (born or conceiv* or deliver* or live* or living)).tw.
17	((birth* or conceiv* or conception* or gestation* or pregnan* or reproducti*) adj2 (achiev* or chance* or natural* or ongoing or outcome* or rate* or spontaneous* or success* or viab*)).tw.
18	or/13-17
19	12 and 18
20	letter.pt. or letter/
21	note.pt.
22	editorial.pt.
23	case report/ or case study/
24	(letter or comment*).ti.
25	or/20-24
26	animal/ not human/
27	nonhuman/
28	exp Animal Experiment/
29	exp Experimental Animal/
30	animal model/
31	exp Rodent/
32	(rat or rats or rodent* or mouse or mice).ti.
33	or/25-32
34	19 not 33
35	limit 34 to english language
36	clinical decision rule/ or biological model/ or statistical model/ or theoretical model/
37	clinical decision making/ or clinical decision support system/ or decision making/ or decision support system/
38	(Hunault* or Lawlor* or McLernon* or Nelson* or Templeton* or van Loendersloot*).tw.
39	(decision* adj2 (clinical* or make or making or model* or support* or system* or technique*)).tw.
40	(cumulati* adj3 (probab* or rate*)).tw.
41	prediction/
42	(predict* or validat*).ti.
43	rule*.tw.
44	(predict* and (model* or outcome* or risk*)).ab.
45	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) and (decision* or identif* or model* or predict* or prognos*)).ti.
46	((characteristic* or criteria or factor* or finding* or history or scor* or variable*) adj3 (decision* or identif* or model* or predict* or prognos*)).ab.
47	decision.tw. and exp statistical model/
48	(prognostic and (characteristic* or criteria or factor* or finding* or history or model* or scor* or variable*)).tw.
49	receiver operating characteristic/
50	(stratification or discrimination or discriminate or c statistic or area under the curve or AUC or calibration or index or indices or algorithm* or logistic regression or multivaria* or nomogram*).tw.
51	or/36-50
52	35 and 51
53	health economics/
54	exp economic evaluation/
55	exp health care cost/
56	exp fee/
57	budget/
58	funding/
59	budget*.ti,ab.
60	cost*.ti.
61	(economic* or pharmaco?economic*).ti.
62	(price* or pricing*).ti,ab.
63	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.

#	Searches
64	(financ* or fee or fees).ti,ab.
65	(value adj2 (money or monetary)).ti,ab.
66	or/53-65
67	52 and 66
68	(conference abstract* or conference review or conference paper or conference proceeding).db,pt,su.
69	67 not 68

1

2 **Database: Health Technology Assessment (HTA) – CRD platform**

3 **Date of last search: 19/07/2023**

ID	Search
1	((artificial* or assist* or intracytoplasm* or intrauter*) and (conception* or inject* or inseminat* or microinject* or reproduct* or sperm* or transfer*))
2	(ART or ICSI or "IC SI" or IUI or IVF or "in vitro fertilisation" or "in vitro fertilization" or "invitro fertilisation" or "invitro fertiliation")
3	((fecund* or fertil* or hypofertil* or infertil* or steril* or subfecund* or subfertil*) and ((intervention* or therap* or treat*) or (coitus or coital or sex* or intercourse*) and (frequen* or regular* or unprotect*)))
4	((active* and surveill*) or (clinical and observ*) or (expect* and (approach* or manag*)) or "without intervention" or "without therapy" or "without treatment" or (wait* and see*) or (watch* and wait*) or untreated)
5	1 or 2 or 3 or 4
6	((birth* or conceiv* or conception* or embryo* or gestation* or pregnan* or reproducti*) and (achiev* or chance* or death* or fail* or live or loss* or natural* or nonviab* or ongoing or outcome* or rate* or spontaneous* or success* or unsuccess* or viab*))
7	5 and 6
8	(((predict* or cumulat* or decision* or likelihood or prognos* or validat*) and (characteristic* or criteria or factor* or finding* or history or model* or outcome* or rate* or risk* or rule* or scor* or variable*)) or stratification or discrimination or discriminate or "c statistic " or "c-statistic " or "area under the curve " or AUC or calibration or index or indices or algorithm* or "logistic regression" or multivaria* or nomogram*)
9	7 and 8 in HTA

4

5 **Database: International Network of Agencies for Health Technology Assessment (INAHTA)**

6 **Date of last search: 19/07/2023**

ID	Search
1	((artificial* or assist* or intracytoplasm* or intrauter*) and (conception* or inject* or inseminat* or microinject* or reproduct* or sperm* or transfer*))
2	(ART or ICSI or "IC SI" or IUI or IVF or "in vitro fertilisation" or "in vitro fertilization" or "invitro fertilisation" or "invitro fertiliation")
3	((fecund* or fertil* or hypofertil* or infertil* or steril* or subfecund* or subfertil*) and ((intervention* or therap* or treat*) or (coitus or coital or sex* or intercourse*) and (frequen* or regular* or unprotect*)))

ID	Search
4	((active* and surveill*) or (clinical and observ*) or (expect* and (approach* or manag*)) or "without intervention" or "without therapy" or "without treatment" or (wait* and see*) or (watch* and wait*) or untreated)
5	1 or 2 or 3 or 4
6	((birth* or conceiv* or conception* or embryo* or gestation* or pregnan* or reproducti*) and (achiev* or chance* or death* or fail* or live or loss* or natural* or nonviab* or ongoing or outcome* or rate* or spontaneous* or success* or unsuccess* or viab*))
7	5 and 6
8	(((predict* or cumulat* or decision* or likelihood or prognos* or validat*) and (characteristic* or criteria or factor* or finding* or history or model* or outcome* or rate* or risk* or rule* or scor* or variable*)) or stratification or discrimination or discriminate or "c statistic " or "c-statistic " or "area under the curve " or AUC or calibration or index or indices or algorithm* or "logistic regression" or multivaria* or nomogram*)
9	7 and 8

1

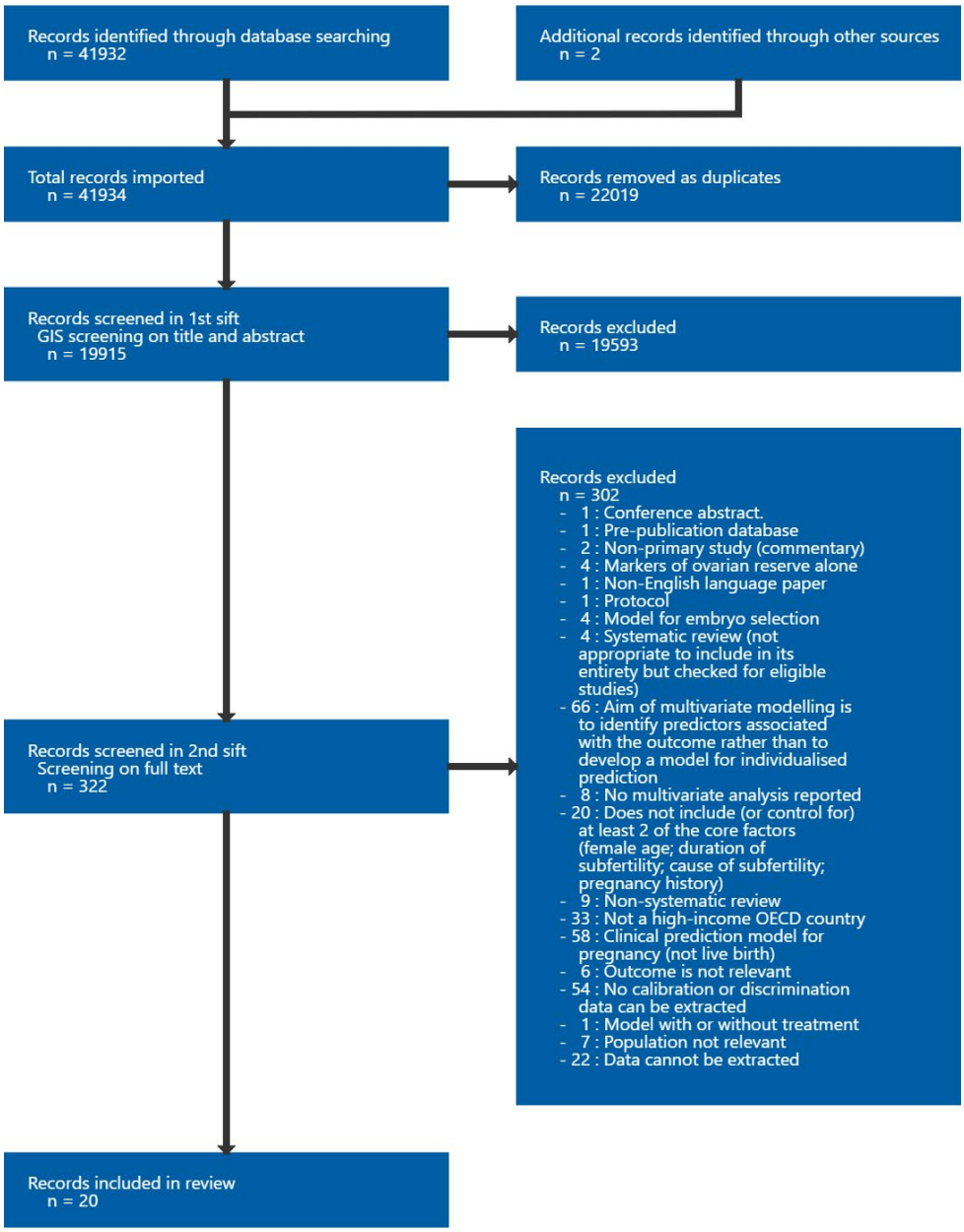
2

Appendix C Clinical prediction model evidence study selection

Study selection for: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

Clinical search

Figure 4: Study selection flowchart



Appendix D Evidence tables

Evidence tables for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

Please refer to Supplement J - Evidence tables for fertility prediction models

Appendix E Forest plots

Forest plots for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

This section includes forest plots only where there was discrimination/calibration data from multiple external validation studies for the same clinical prediction model. Model performance data from single studies is not presented here; the quality assessment for such outcomes is provided in the GRADE profiles in appendix F.

Figure 5: Meta-analysis C-statistic estimates for the Hunault (2004) model

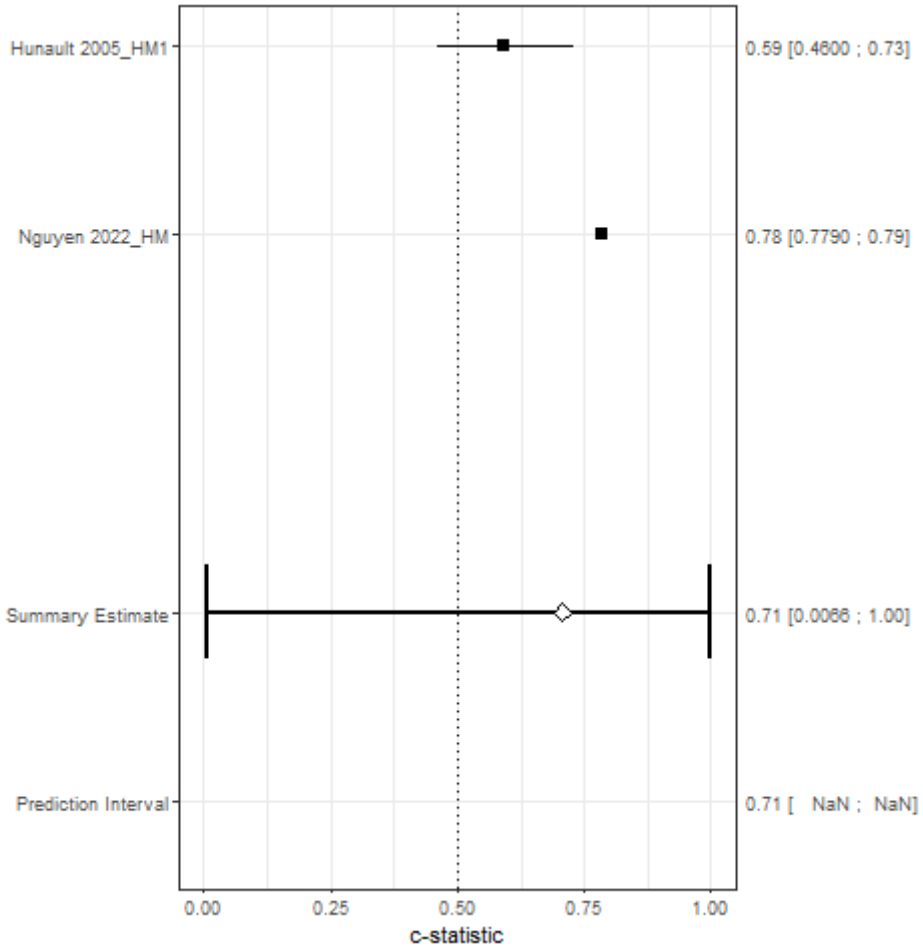


Figure 6: Meta-analysis O:E ratio estimates for the Hunault (2004) model

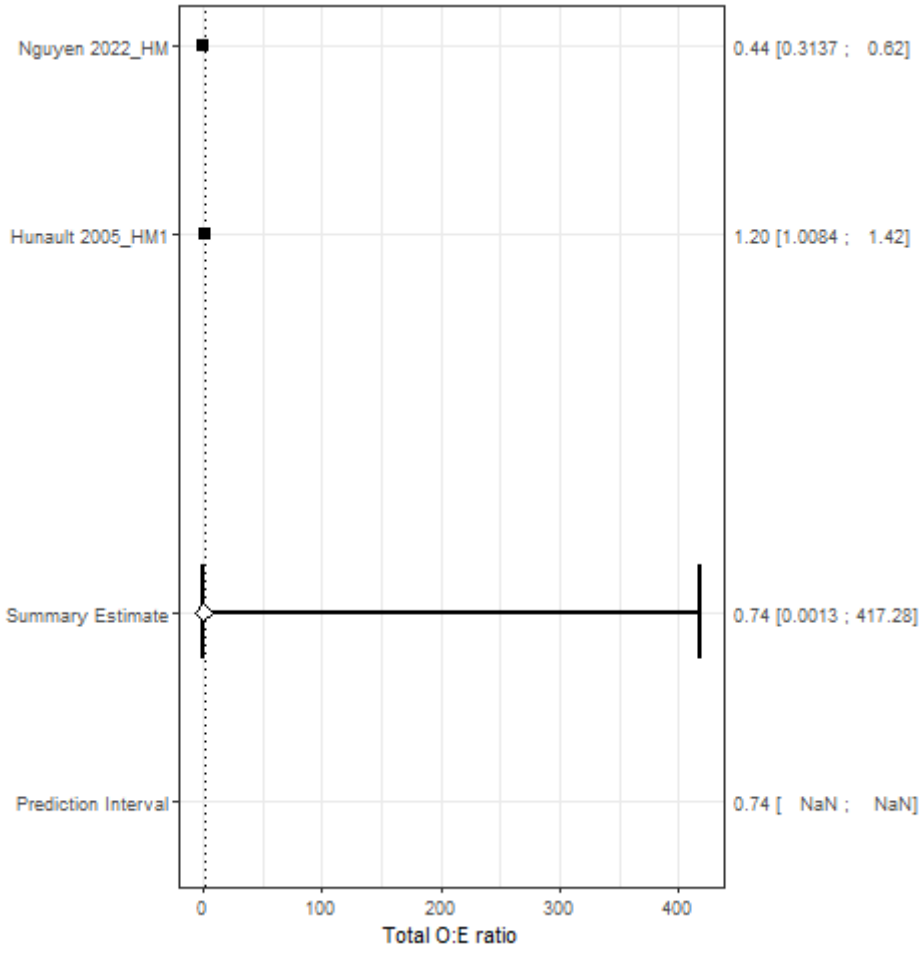


Figure 7: Meta-analysis C-statistic estimates for the Templeton (1996) model

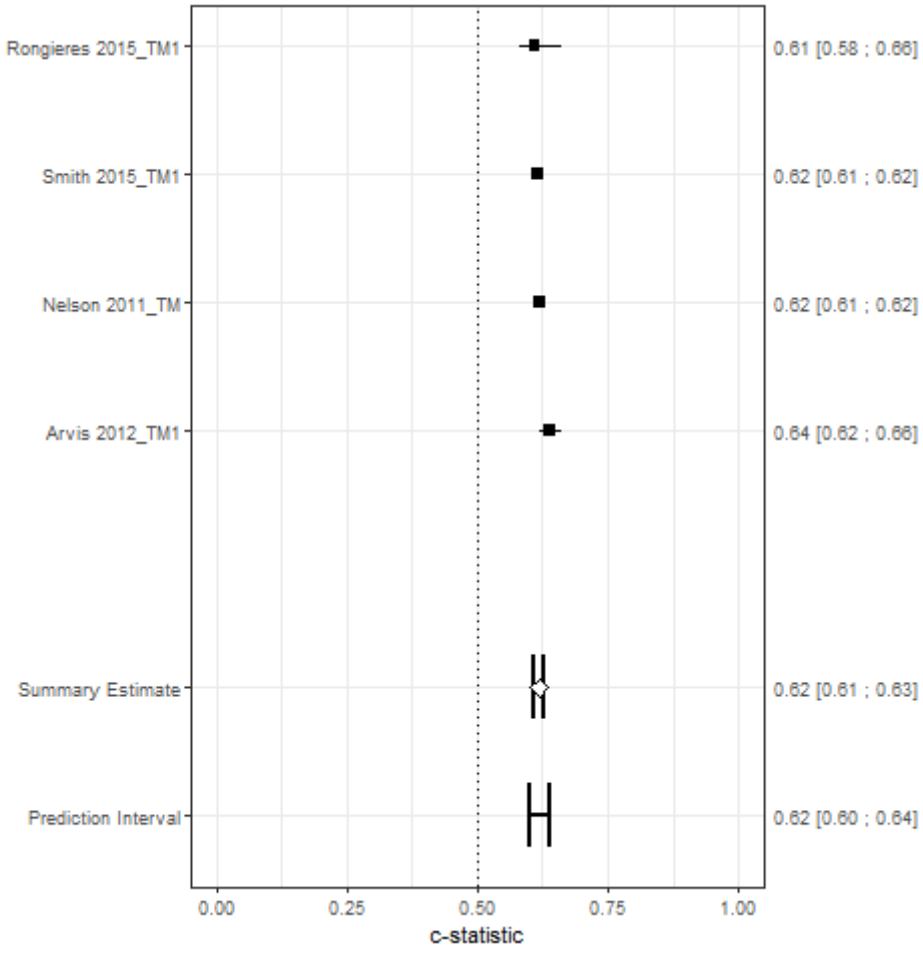


Figure 8: Meta-analysis C-statistic estimates for the Templeton (1996) model, with centre-specific fitting

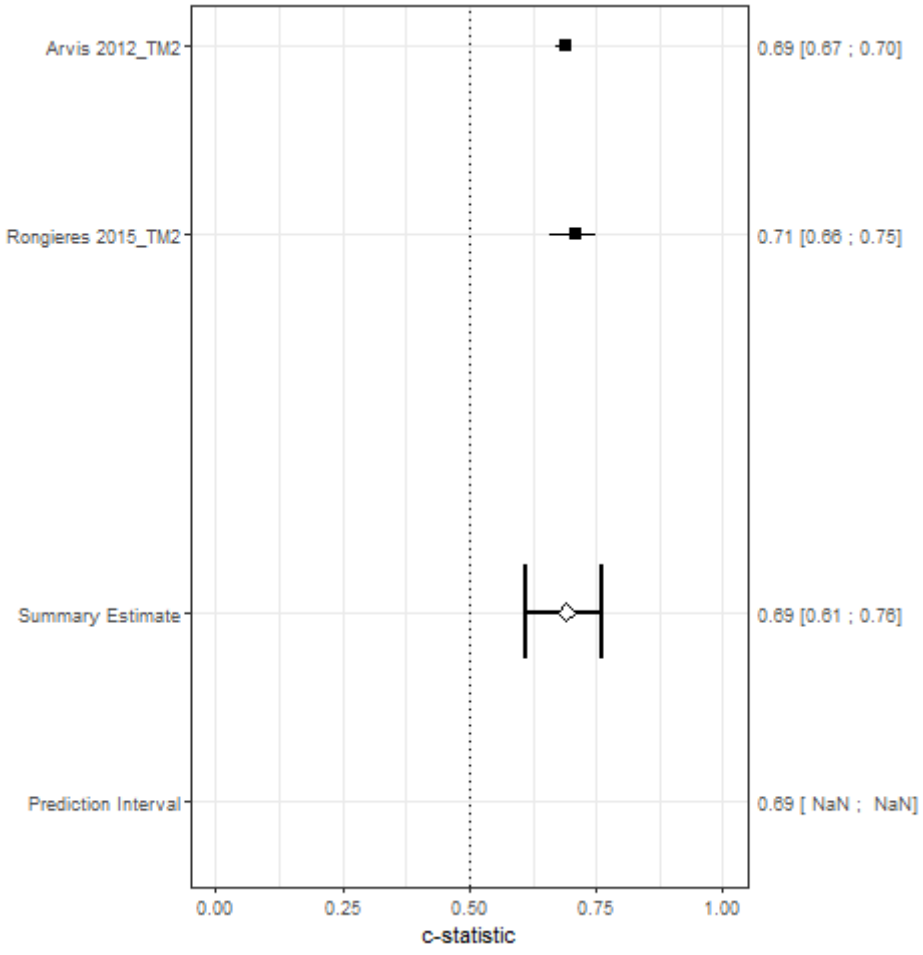


Figure 9: Meta-analysis C-statistic estimates for the McLernon (2016) pre-treatment model

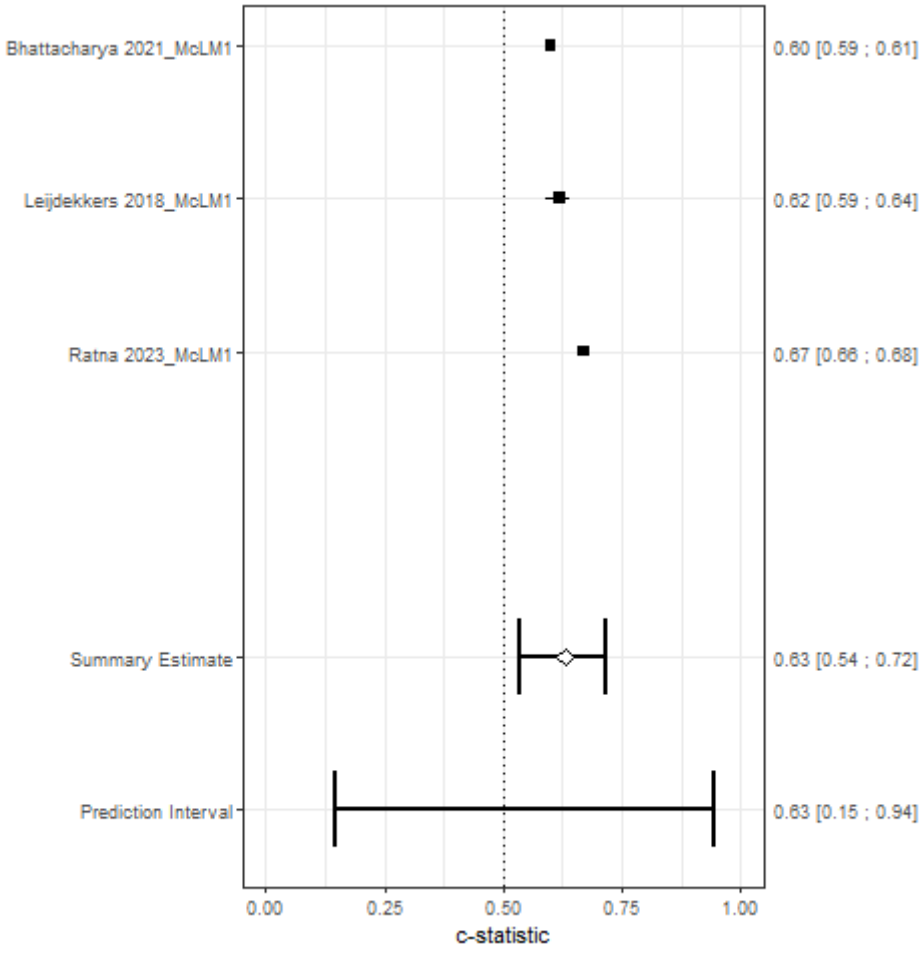


Figure 10: Meta-analysis C-statistic estimates for the McLernon (2016) post-treatment model

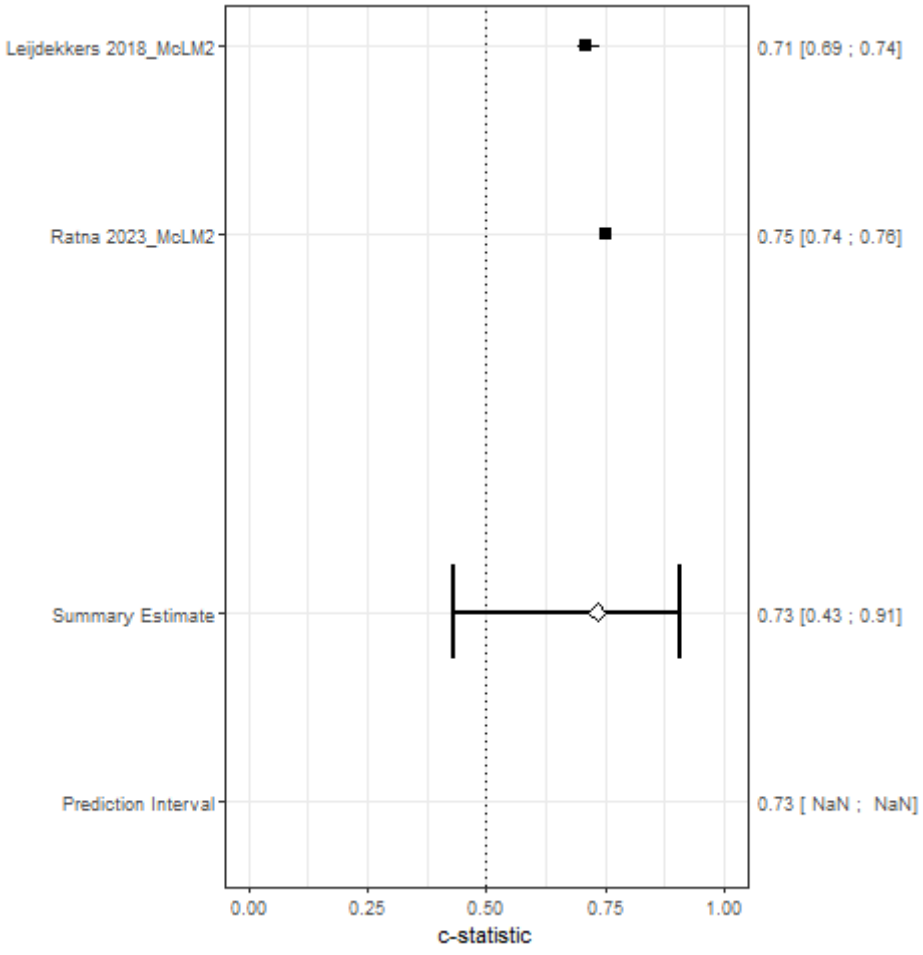


Figure 11: Meta-analysis C-statistic estimates for the van Loendersloot (2013) model

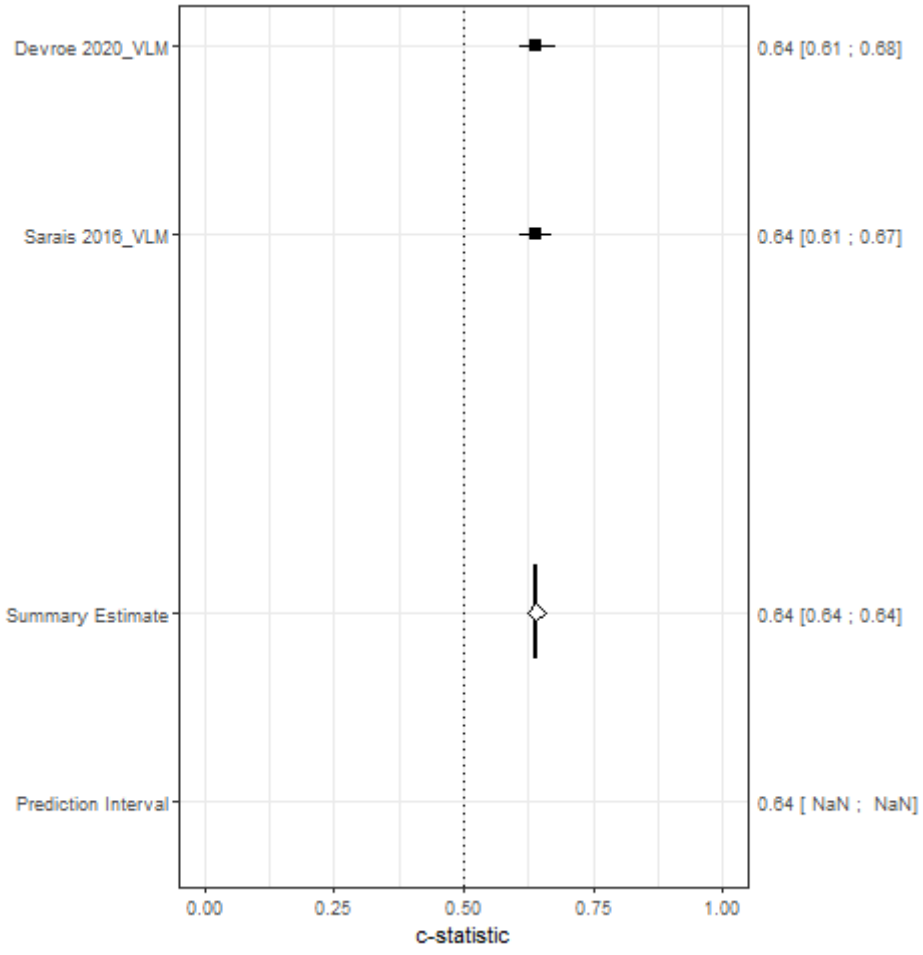
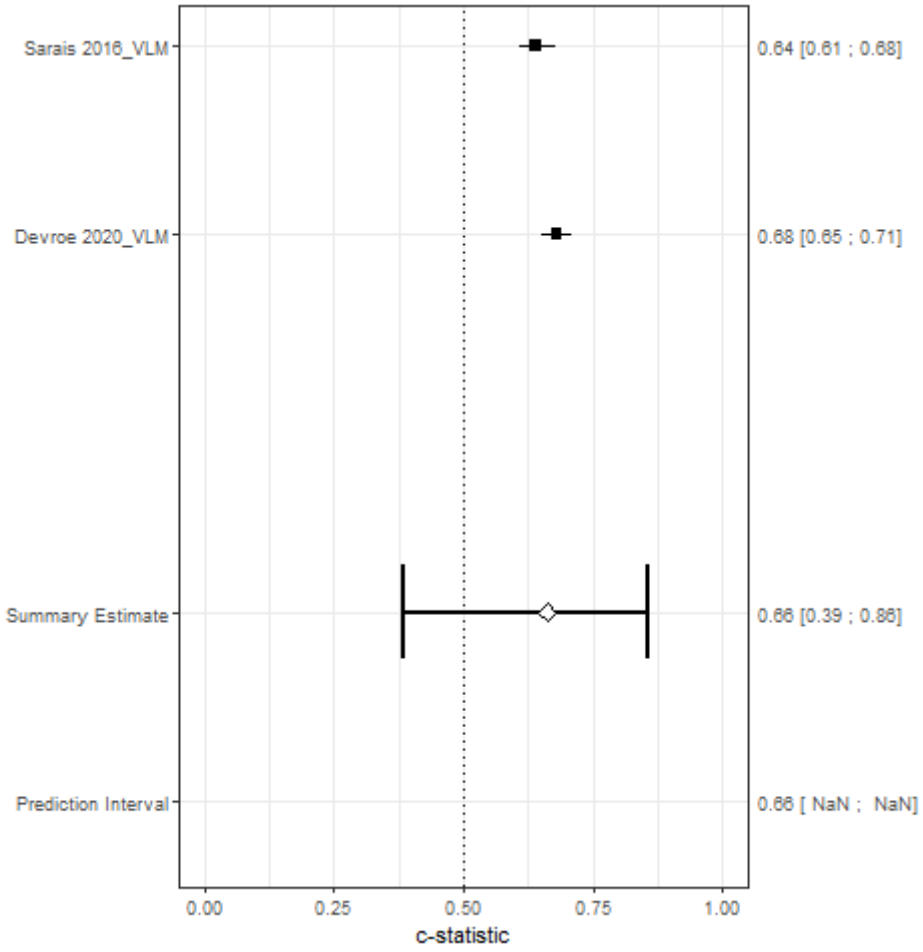


Figure 12: Meta-analysis C-statistic estimates for the van Loendersloot (2013) model, with centre-specific fitting



Appendix F GRADE tables

GRADE tables for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

Table 5: Evidence profile for clinical prediction models for expectant management

Quality assessment								Model performance		Quality	Importance
No of studies	Model/ predictors	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	No of participants/ cycles	C-statistic (95% CI)	Number of observed and expected live births, and O:E ratio		
External validation											
2[1]	HM (without PCT)	very serious[2]	not estimable[3]	no serious indirectness	very serious[4]	none	N participants: 627 N cycles: NR	0.71 (0.007 to 1.00)	Observed: 121 Expected: 144 O:E ratio (95% CI): 0.74 (0.001 to 417)	VERY LOW	CRITICAL
1 (Hunault 2005)	HM (with PCT)	very serious[2]	no serious inconsistency	no serious indirectness	serious[5]	none	N participants: 302 N cycles: NR	0.63 (0.51 to 0.75)	Observed: 91 Expected: 95 O:E ratio (95% CI): 0.96 (NR)	VERY LOW	CRITICAL
Model development											
1 (Collins 1995)	•Duration of subfertility (≤36 months) •Female age (≤30 years) •Pregnancy history	very serious[2]	no serious inconsistency	no serious indirectness	serious[5]	none	N participants: 2198 N cycles: NR	0.59 (0.56 to 0.63)	Observed: 263 Expected: NR O:E ratio: NR	VERY LOW	CRITICAL

	(secondary infertility) •Cause of subfertility (tubal; endometriosis) •Male factor (oligospermia or azoospermia)										
1 (Nguyen 2022)	•Duration of subfertility (continuous) •Female age (continuous) •Pregnancy history (secondary infertility; primary infertility) •Sperm motility (continuous) •Serum AMH (ng/ml)	very serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	N participants: 325 N cycles: NR	0.797 (0.792 to 0.802)	Observed: 30 Expected: NR O:E ratio: NR	LOW	CRITICAL
1 (Snick 1997)	Model 1: •Duration of subfertility (<24 months) •Cause of subfertility (tubal; ovulatory) •Post-coital test (abnormal)	very serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	N participants: 726 N cycles: NR	0.79 (0.75 to 0.83)	Observed: 201 Expected: NR O:E ratio: NR	LOW	CRITICAL

1 (Snick 1997)	Model 2: •Duration of subfertility (<24 months) •Female age (<30 years) •Pregnancy history (secondary infertility) •Cause of subfertility (tubal; ovulatory) •Male factor (WHO semen abnormality)	very serious[2]	no serious inconsistency	no serious indirectness	no serious imprecision[6]	none	N participants: 726 N cycles: NR	Development dataset: 0.76 (0.72 to 0.8) Validation dataset: 0.67 (0.63 to 0.7)	Observed: 201 Expected: NR O:E ratio: NR	LOW	CRITICAL

AMH: anti-Müllerian hormone; C-statistic: concordance statistic; CI: confidence interval; HM: Hunault (2004) model; NR: not reported; O:E: observed:expected; PCT: post-coital test; PROBAST: Prediction model Risk Of Bias ASsessment Tool
[1]Hunault 2005; Nguyen 2022
[2] Very serious risk of bias in the evidence contributing to the outcome as per PROBAST checklist
[3] Prediction interval could not be calculated as <3 studies
[4] 95% CI crosses 2 clinical decision making thresholds
[5] 95% CI crosses 1 clinical decision making threshold
[6] No serious imprecision based on the validation dataset which is the higher level of evidence

Table 6: Evidence profile for clinical prediction models for intrauterine insemination (IUI)

Quality assessment								Model performance		Quality	Importance
No of studies	**Model/ predictors**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**No of participants/ cycles**	**C-statistic (95% CI)**	**Number of observed and expected live births, and O:E ratio**		
Model development											

1 (Hansen 2016)	•Duration of subfertility (continuous; in months) •Female age (continuous) •Pregnancy history (prior loss yes/no) •Ovarian stimulation agent (clomifene citrate; letrozole; gonadotropin) •Income (<$50,000; ≥$50,000)	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	N participants: 900 N cycles: 2572	0.65 (0.61 to 0.69)	Observed: 223 Expected: NR O:E ratio (95% CI): NR	LOW	CRITICAL

C-statistic: concordance statistic; CI: confidence interval; NR: not reported; O:E: observed:expected; PROBAST: Prediction model Risk Of Bias ASsessment Tool
[1] Very serious risk of bias in the evidence contributing to the outcome as per PROBAST checklist

Table 7: Evidence profile for clinical prediction models for in vitro fertilisation (IVF)

Quality assessment								Model performance		Quality	Importance
No of studies	**Model/ predictors**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**No of participants/ cycles**	**C-statistic (95% CI)**	**Number of observed and expected live births, and O:E ratio**		
External validation											
4[1]	TM	very serious[2]	no serious inconsistency	serious[3]	no serious imprecision	none	N participants: NR N cycles: 288594	0.62 (0.61 to 0.63)	Observed: 69476 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL

2[4]	TM (centre-specific fitting)	very serious[2]	not estimable[5]	serious[3]	serious[6]	none	N participants: NR N cycles: 13616	0.69 (0.61 to 0.76)	Observed: 2399 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL
1 (Rongieres 2015)	TMA	very serious[2]	no serious inconsistency	serious[3]	serious[6]	none	N participants: NR N cycles: 715	0.76 (0.71 to 0.80)	Observed: 108 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL
1 (Smith 2015)	NLM	serious[7]	no serious inconsistency	serious[3]	no serious imprecision	none	N participants: NR N cycles: 130960	0.628 (0.625 to 0.631)	Observed: 33526 Expected: 30383 O:E ratio (95% CI): 0.905 (0.896 to 0.913)	LOW	CRITICAL
3[8]	McLM1 (pre-treatment)	serious[7]	very serious[9]	serious[3]	serious[6]	none	N participants: 102135 N cycles: NR	0.63 (0.54 to 0.72)	Observed: NR Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL
2[10]	McLM1 (post-treatment)	serious[7]	not estimable[5]	serious[3]	very serious[11]	none	N participants: 92546 N cycles: 147615	0.73 (0.43 to 0.91)	Observed: 41426 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL
2[12]	VLM	very serious[2]	not estimable[5]	very serious[13]	no serious imprecision	none	N participants: 1363 N cycles: 2053	0.64 (0.64 to 0.64)	Observed: 555 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL

2[12]	VLM (centre-specific fitting)	very serious[2]	not estimable[5]	very serious[13]	very serious[11]	none	N participants: 1363 N cycles: 2053	0.66 (0.39 to 0.86)	Observed: 555 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL
Model development											
1 (Arvis 2012)	Model development (TM + additional variables): •Duration of subfertility 1 year; 4 years; 7 years; 13 years) •Female age (quadratic and cubic polynomial components of age) •Female BMI (>26 or <18) •Pregnancy history (previous LB with IVF; previous non-LB with IVF; previous LB not by IVF; previous non-LB not by IVF) •Previous IVF treatment outcome (number of previous unsuccessful IVF attempts)	very serious[2]	no serious inconsistency	serious[3]	no serious imprecision	none	N participants: NR N cycles: 12901	0.71 (0.68 to 0.74)	Observed: 2291 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL

	•Cause of subfertility (tubal factor) •Markers of ovarian reserve (FSH>10) •Smoking (smokes or smoked in the past) •Year (from 2011)										
1 (Balachandren 2020)	•Female age (36-37; 38-39; 40-42) •Markers of ovarian reserve (FSH>12; AMH>8.5)	very serious[2]	no serious inconsistency	very serious[13]	no serious imprecision	none	N participants: NR N cycles: 516	0.68 (0.63 to 0.73)	Observed: 357 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL
1 (Devroe 2020)	Model development (VLM + additional variables): •Duration of subfertility (continuous; 5 years) •Female age (continuous) •Pregnancy history (previous delivery yes/no) •Previous IVF treatment outcome (number of previous failed	very serious[2]	no serious inconsistency	very serious[13]	no serious imprecision	none	N participants: 591 N cycles: 1281	0.71 (0.68 to 0.75)	Observed: 344 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL

	IVF/ICSI cycles: 0; 1; 2; 3; 4; 5) •Markers of ovarian reserve (FSH continuous; FSH≤10) •Male factor infertility (yes/no) •Embryo quality (morphological score of all embryos day 3 in previous cycle) •Embryo grade (≥1 8-cell embryo on day 3 in previous cycle; ≥1 morula on day 3 in previous cycle; ≥1 8-cell embryo on day 3 in current cycle; ≥1 morula on day 3 in current cycle) •Number of oocytes retrieved (number of embryos after oocyte retrieval in previous cycle; number with ≥10 embryos after oocyte retrieval in previous cycle; number of embryos										

	after oocyte retrieval in current cycle; number with ≥10 embryos after oocyte retrieval in current cycle) •Endometriosis (yes/no)										
1 (Dhillon 2016)	•Duration of subfertility (0-4 years; ≥5 years) •Female age (≤36 years; >36 years) •BMI (continuous) •Pregnancy history (previous live birth; previous miscarriage) •Cause of subfertility (male factor; tubal factor; anovulation; unexplained; other) •Markers of ovarian reserve (AFC continuous; AFC squared) •Male factor infertility (yes/no) •Ethnicity (white; Asian; black; Chinese;	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	N participants: 9915 N cycles: 9915	0.62 (0.60 to 0.64)	Observed: 863 Expected: NR O:E ratio (95% CI): NR	HIGH	CRITICAL

	other; not stated; mixed)										
1 (Hamdine 2015)	•Female age (continuous) •Pregnancy history (primary infertility; secondary infertility) •Previous IVF treatment outcome (number of previous ART treatments) •Markers of ovarian reserve (AMH continuous)	very serious[2]	no serious inconsistency	no serious indirectness	serious[6]	none	N participants: 487 N cycles: 1363	0.59 (0.56 to 0.63)	Observed: 245 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL
1 (La Marca 2021)	Model 1 (pre-treatment): •Duration of subfertility (<1 year; 1-3; 4-6; 7-9; 10-12; >12) •Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) •Pregnancy history (female primary infertility yes/no) •Cause of subfertility (unknown; low sperm count only; ovulatory only; tubal	very serious[2]	no serious inconsistency	very serious[13]	no serious imprecision	none	N participants: 57699 N cycles: 57699	0.61 (0.60 to 0.61)	Observed: NR Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL

	disease only; other)										
1 (La Marca 2021)	Model 2 (current cycle variable added to pre-treatment model): •Duration of subfertility (<1 year; 1-3; 4-6; 7-9; 10-12; >12) •Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) •Pregnancy history (female primary infertility yes/no) •Cause of subfertility (unknown; low sperm count only; ovulatory only; tubal disease only; other) •Number of oocytes retrieved (1-4; 5-9; 10-14; 15-19; ≥20)	very serious[2]	no serious inconsistency	very serious[13]	no serious imprecision	none	N participants: 52960 N cycles: 52960	0.67 (0.66 to 0.67)	Observed: NR Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL
1 (La Marca 2021)	Model 3 (current cycle variables added to pre-treatment model):	very serious[2]	no serious inconsistency	very serious[13]	no serious imprecision	none	N participants: 50870 N cycles: 50870	0.65 (0.64 to 0.65)	Observed: NR Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL

	•Duration of subfertility (<1 year; 1-3; 4-6; 7-9; 10-12; >12) •Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) •Pregnancy history (female primary infertility yes/no) •Cause of subfertility (unknown; low sperm count only; ovulatory only; tubal disease only; other) •Number of oocytes retrieved (1-4; 5-9; 10-14; 15-19; ≥20) •Number of embryos created (1-4; 5-9; 10-14; 15-19; ≥20)										
1 (Leijdekkers 2018)	Model development pre-treatment (McLM1 + additional variables): •Duration of subfertility (continuous, range of	serious[7]	no serious inconsistency	no serious indirectness	no serious imprecision	none	N participants: 1511 N cycles: 2881	0.66 (0.64 to 0.68)	Observed: 774 Expected: NR O:E ratio (95% CI): NR	MODERATE	CRITICAL

	possible values 0-21) •Female age (continuous, range of possible values 18-50) •BMI (weight, kg) •Pregnancy history (previous pregnancy in couple yes/no) •Cause of subfertility (tubal factor; anovulation; male factor; unexplained) •Markers of ovarian reserve (AMH, ng/l; AFC, 2-10mm) •Male factor infertility (yes/no) •Number of cycles (cycle number) •Treatment type (IVF or ICSI) •Year of first egg retrieval (as restricted cubic spline)										
1 (Leijdekkers 2018)	Model development post-treatment (McLM2 +	serious[7]	no serious inconsistency	serious[3]	no serious imprecision	none	N participants: 1511	0.71 (0.69 to 0.73)	Observed: 774 Expected: NR	LOW	CRITICAL

	additional variables): •Duration of subfertility (continuous, range of possible values 0-21) •Female age (continuous, range of possible values 18-50) •Pregnancy history (primary infertility of couple yes/no) •Cause of subfertility (tubal factor yes/no) •Markers of ovarian reserve (AMH, ng/l; AFC, 2-10mm) •Embryo grade (1st fresh embryo transfer: no embryos transferred; single cleavage stage; single blastocyst stage; double cleavage stage; double blastocyst stage; triple cleavage stage; triple blastocyst stage)						N cycles: 2881		O:E ratio (95% CI): NR		

	•Number of eggs collected in 1st cycle (range of possible values 1-28) •Fresh or frozen embryo transfer (embryos frozen in 1st cycle yes/no) •Number of cycles (cycle number) •Treatment type (IVF or ICSI in 1st cycle) •Year of first complete cycle (as restricted cubic spline)										
1 (McLernon 2016)	Model 1 (pre-treatment): •Duration of subfertility (continuous) •Female age (continuous, as restricted cubic spline) •Pregnancy history (previous pregnancy in couple yes/no) •Cause of subfertility (tubal factor; anovulation;	serious[7]	no serious inconsistency	no serious indirectness	no serious imprecision	none	N participants: 113873 N cycles: 184269	0.73 (0.72 to 0.74)	Observed: 48925 Expected: NR O:E ratio (95% CI): NR	MODERATE	CRITICAL

	male factor; unexplained) •Male factor infertility (yes/no) •Number of cycles (cycle number) •Treatment type (IVF or ICSI) •Year of first egg retrieval (as restricted cubic spline)										
1 (McLernon 2016)	Model 2 (post-treatment): •Duration of subfertility (continuous) •Female age (continuous, as restricted cubic spline) •Pregnancy history (primary infertility of couple yes/no) •Cause of subfertility (tubal factor yes/no) •Embryo grade (1st fresh embryo transfer: no embryos transferred; single cleavage stage; single blastocyst	serious[7]	no serious inconsistency	serious[3]	no serious imprecision	none	N participants: 113873 N cycles: 184269	0.72 (0.71 to 0.73)	Observed: 48925 Expected: NR O:E ratio (95% CI): NR	LOW	CRITICAL

	stage; double cleavage stage; double blastocyst stage; triple cleavage stage; triple blastocyst stage) •Number of eggs collected in 1st cycle (as restricted cubic spline) •Fresh or frozen embryo transfer (embryos frozen in 1st cycle yes/no) •Number of cycles (cycle number) •Treatment type (IVF or ICSI in 1st cycle) •Year of first complete cycle (as restricted cubic spline)										
1 (Meijerink 2016)	•Female age (age squared) •[Cause of subfertility not included as factor but inclusion criteria male factor] •Sperm motility (oocytes injected with	serious[7]	no serious inconsistency	very serious[13]	no serious imprecision	none	N participants: 289 N cycles: 553	Development: 0.62 (0.58 to 0.68) Validation: 0.67 (0.62 to 0.72)	Observed: 113 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL

	motile; immotile; both motile and immotile) •Male factor infertility (suspected diagnosis before sperm retrieval: obstructive azoospermia; non-obstructive azoospermia) •Number of cycles (cycle number) •Male hormones level (LH continuous; testosterone continuous)										
1 (Nelson 2011)	Model development (TM + additional variables): •Duration of subfertility (<1 year; 1-3 years; 4-6 years; 7-9 years; 9-12 years; >12 years) •Female age (18-34; 35-37; 38-39; 40-42; 43-44; 45-50) •Pregnancy history (previous LB with IVF;	very serious[2]	no serious inconsistency	serious[3]	no serious imprecision	none	N participants: NR N cycles: 144018	0.63 (0.62 to 0.64)	Observed: 33524 Expected: NR O:E ratio (95% CI): NR	VERY LOW	CRITICAL

	previous non-LB with IVF; previous LB not by IVF; previous non-LB not by IVF) •Previous IVF treatment outcome (number of previous unsuccessful IVF attempts) •Cause of subfertility (unknown; tubal only; anovulatory only; endometriosis only; cervical only; male only; combination known causes) •Number of cycles (cycle number 1,2,3+) •Treatment type (IVF or ICSI) •Hormonal preparation (antioestrogen; gonadotropin; hormone replacement) •Source of egg (donor; participant)										
1 (Rongieres 2015)	Model development (TM +/- variables):	very serious[2]	no serious inconsistency	serious[3]	serious[6]	none	N participants: NR	0.75 (0.71 to 0.80)	Observed: 108 Expected: NR	VERY LOW	CRITICAL

	•Pregnancy history (previous LB with IVF) •Markers of ovarian reserve (AMH continuous) •Year (from 2011)						N cycles: 715		O:E ratio (95% CI): NR		

AFC: antral follicle count; AMH: anti-Müllerian hormone; BMI: body mass index; C-statistic: concordance statistic; CI: confidence interval; FSH: follicle stimulating hormone; ICSI: intracytoplasmic sperm injection; IVF: in vitro fertilisation; LB: live birth; LH: luteinizing hormone; McLM: McLernon model; ng/ml: nanograms per millilitre; NLM: Nelson Lawler model; NR: not reported; O:E: observed:expected; PROBAST: Prediction model Risk Of Bias ASsessment Tool; TM: Templeton (1996) model; TMA: Templeton-Arvis model; VLM: van Loendersloot model

[1]Arvis 2012; Nelson 2011; Rongieres 2015; Smith 2015

[2] Very serious risk of bias in the evidence contributing to the outcome as per PROBAST checklist

[3] Serious concern for applicability in the evidence contributing to the outcome as per PROBAST checklist

[4]Arvis 2012; Rongieres 2015

[5] Prediction interval could not be calculated as <3 studies

[6] 95% CI crosses 1 clinical decision making threshold

[7] Serious risk of bias in the evidence contributing to the outcome as per PROBAST checklist

[8]Bhattacharya 2021; Leijdekkers 2018; Ratna 2023

[9] Very serious heterogeneity indicated by prediction interval

[10] Leijdekkers 2018; Ratna 2023

[11] 95% CI crosses 2 clinical decision making thresholds

[12]Devroe 2020; Sarais 2016

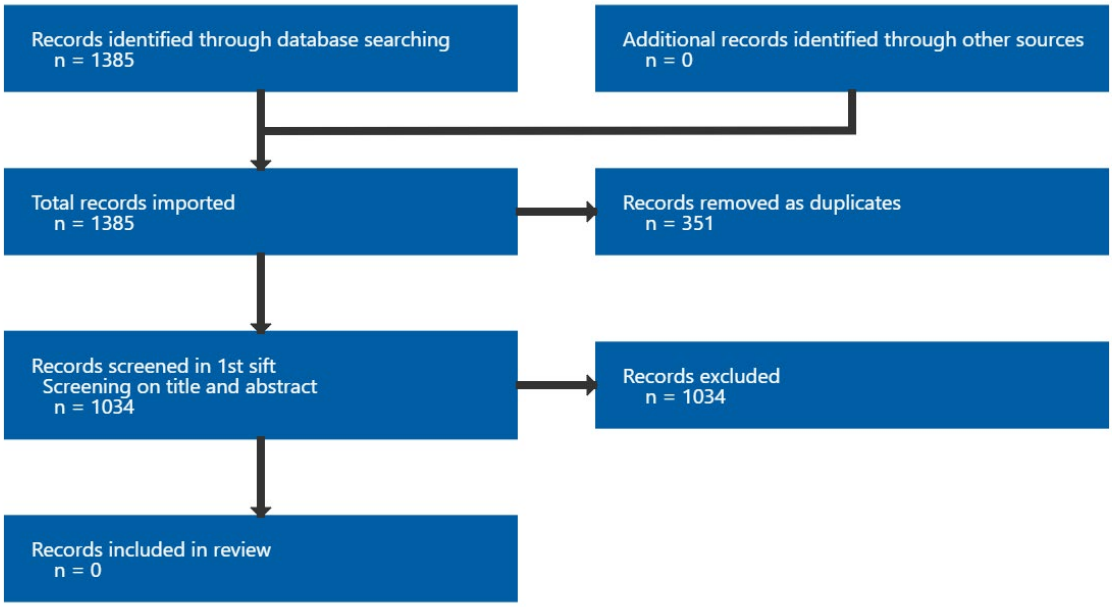
[13] Very serious concern for applicability in the evidence contributing to the outcome as per PROBAST checklist

Appendix G Economic evidence study selection

Study selection for: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

No evidence was identified which was applicable to this review question

Figure 13: Study selection flow chart



Appendix H Economic evidence tables

Economic evidence tables for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

No evidence was identified which was applicable to this review question.

Appendix I Economic model

Economic model for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

Cost-utility analysis of IVF provision on the NHS

Introduction

The NICE 2013 guideline (CG156) made recommendations on access criteria to IVF treatment on the NHS based on an economic evaluation. This analysis relied on the use of prediction models to estimate the treatment effectiveness of IVF, but we were aware that newer prediction models have been developed subsequently. Therefore, a new economic model was developed for this guideline to utilise this newer prediction model evidence.

Methods

Setting and population

The model setting was for the NHS and the population was people with health-related fertility problems. The model population comprised women of reproductive age, from 20-45 years of age. Not achieving a live birth from spontaneous conception after 2 years of trying was accepted as establishing an unexplained health-related fertility problem in those without an identified cause of subfertility. Therefore, the lower age limit in the model was derived from the beginning of adulthood plus 2 years of trying to conceive spontaneously. The upper age limit was accepted by the committee as being a reasonable approximation for the end of a woman's reproductive life, whilst recognising that spontaneous conception can occur in older women.

The model assesses the cost-effectiveness of IVF for the following causes of subfertility.

- Unknown
- Tubal
- Anovulation
- Mild endometriosis
- Severe endometriosis
- Male factor
- Cervical
- Combined

Time horizon

To estimate cumulative live births the time horizon of the model was from the age of starting treatment to the end of the woman's reproductive life. This was adopted to reflect that expectant management is not time limited and that spontaneous conception can still occur following unsuccessful fertility treatment. However, it was assumed that improvements in health state utility from a live birth would be lifelong and therefore a longer lifetime horizon was used to estimate the Quality Adjusted Life Year (QALY) gain from a live birth. This was calculated from the age at giving birth to death, based on the remaining life expectancy for a woman of that age.

The model assumed a 2-month interval spacing between IVF cycles, but this was varied in a sensitivity analysis.

Model structure

The health economic model was developed in Microsoft Excel® to compare the cost effectiveness of 1 to 6 cycles of IVF. The model was restricted to a maximum of 6 cycles, reflecting the upper limit of cycles in prediction models for IVF utilised in this analysis.

The model included the following treatment strategies:

1. Expectant management (EM) for the remainder of the woman's reproductive life without IVF (no IVF)

2. One cycle of IVF, followed by EM for the remainder of the woman's reproductive life if 1 full cycle of IVF was unsuccessful (IVF1)

3. Up to 2 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 2 cycles of IVF were unsuccessful (IVF2)

4. Up to 3 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 3 cycles of IVF were unsuccessful (IVF3).

5. Up to 4 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 4 cycles of IVF were unsuccessful (IVF4).

6. Up to 5 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 5 cycles of IVF were unsuccessful (IVF5).

7. Up to 6 cycles of IVF, followed by EM for the remainder of the woman's reproductive life if the 6 cycles of IVF were unsuccessful (IVF6).

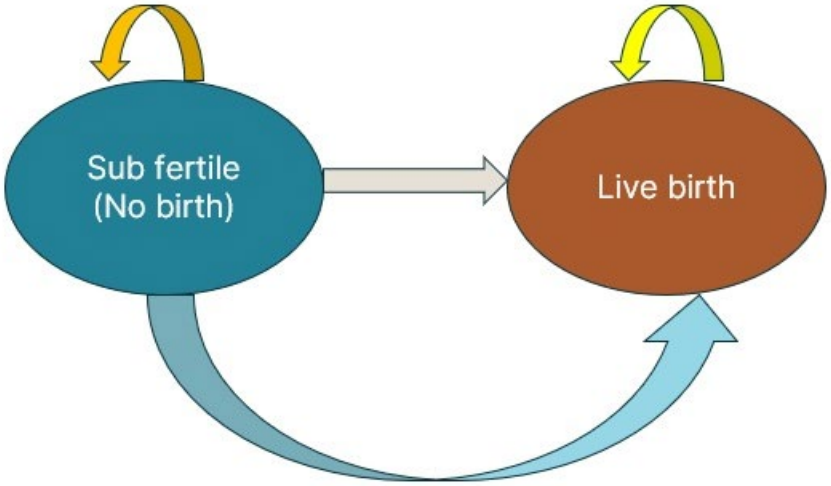
Modelling approach

The model utilised a simple Markov approach as shown in Figure 14. The model used a Markov cycle of 1-month to approximate the menstrual cycle. In each cycle women who had not already had a live birth would either remain in a "no birth" health state or transition to the "live birth" state if they had a successful outcome from fertility treatment or a birth from spontaneous conception if treatment was completed. "Live birth" was an absorbing state with any future fertility not considered in the analysis.

The model accounted for the possibility of live birth from expectant management after completion of active treatment for the woman's remaining reproductive life, which was estimated using a published prediction model (van Eekelen 2017).

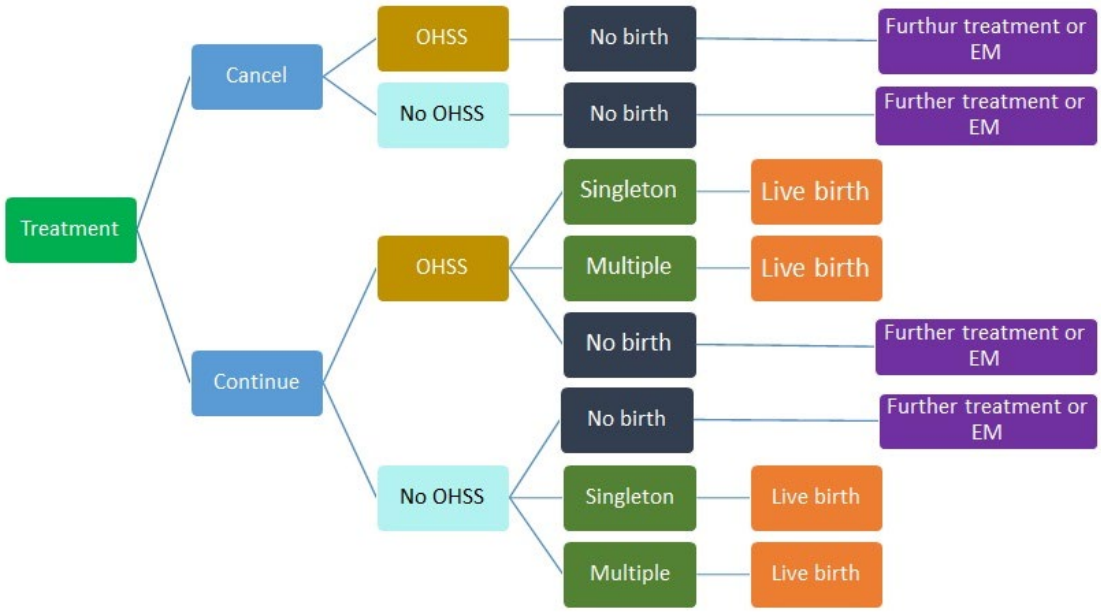
The decision tree for fertility treatment for the differing model outcomes and events associated with that treatment is illustrated in Figure 15.

Figure 14: Markov schematic to assess fertility treatments across a woman's reproductive life cycle



1

Figure 15: Decision tree illustrating the outcomes of IVF



2 ***Clinical outcomes***

3 The clinical outcomes incorporated into the model were:

- Live birth
- Singleton/multiple birth
- Ovarian hyperstimulation syndrome (OHSS)

The purpose of fertility treatment is to increase live birth rates compared to those that would be achieved in the absence of treatment, and therefore that is clearly the most critical clinical outcome in the model. However, fertility treatment is also associated with higher rates of multiple birth and can lead to OHSS. Multiple pregnancy from IVF is not universally considered an undesirable outcome of fertility treatment (Gleicher, 2009) especially where more than one child is wanted to "complete the family". However, as multiple pregnancy is linked to much higher rates of preterm birth it was included within the model in order to capture its higher costs and to reflect that policy makers have worked to reduce the multiple birth rate from IVF (Our campaign to reduce multiple births | HFEA). OHSS is known as important complication of fertility treatment which can have implications for health-related quality of life and costs.

Prediction models

Prediction models were used in this evaluation in order to estimate the effective of IVF treatment relative to no treatment or EM.

Four prediction models were identified for use in the health economic model:

i. van Eekelen (2017) – base case analysis prediction model for live birth arising from spontaneous conception
ii. OPIS pre-IVF (McLernon 2016, Ratna 2023) – base case analysis prediction model for live birth resulting from IVF
iii. Hunault (2004) – sensitivity analysis prediction model for live birth arising from spontaneous conception
iv. IVF Predict (Nelson 2011) – sensitivity analysis prediction model for live birth resulting from IVF

The Hunault prediction model and IVF Predict were the prediction models used for the economic analysis undertaken in the previous guideline. Both models were identified in our systematic review of the prediction model literature and have been retained for the purposes of sensitivity analysis. However, 2 newer prediction models were preferred for the base case analysis.

Unlike the Hunault prediction model, the van Eekelen prediction model is a dynamic prediction model. Such dynamic prediction models are better suited to this analysis as they allow repeated predictions to be made over different periods of time giving estimates of cumulative live birth rates. Hunault is not a dynamic prediction model, and its output represent a prediction of live birth occurring in the next year. Whilst it is possible to generate cumulative live birth rates using the Hunault model by adjusting the model predictors as they change over time, the model was not designed or validated to predict cumulative live birth rates over time. Furthermore, the previous NICE fertility guideline CG156 noted, as a limitation of the Hunault model, that it was based on cohorts where the average age of the population is younger than would be observed in the population covered by this model. Based on their clinical knowledge and experience, the previous guideline committee believed that Hunault overestimated spontaneous conception leading to live birth, especially in older age groups.

The van Eekelen prediction model was excluded from our systematic review because it used ongoing pregnancy at a gestational age of at least 12 weeks rather than live birth as an outcome. Our review protocol stipulated live birth as an outcome as we did not want to include studies which used clinical pregnancy instead (a clinical pregnancy is a pregnancy confirmed by ultrasound). Clinical pregnancy is a common surrogate for live birth in fertility research as it requires less follow-up, but for the purpose of this analysis we wanted a prediction of live birth as that is true outcome of interest. However, ongoing pregnancy at a gestational age of at least 12 weeks is a far better proxy for live birth than clinical pregnancy as it is estimated that around 95% of such pregnancies would result in a live birth (Braakhekke 2014) compared to around 82% from clinical pregnancies resulting from assisted reproduction (CDC 2021). Therefore, given its other advantages over the Hunault model, we believed that the van Eekelen was the best model to use for the estimation of spontaneous conception to live birth. It should be noted that for severe endometriosis, tubal, anovulatory and cervical causes of infertility it was assumed that there was a zero probability of spontaneous conception leading to live birth and therefore no EM prediction was required for the analysis of these causes.

The OPIS pre-IVF model, which was included in our systematic review, was preferred to IVF predict for prediction of live birth arising from IVF. This was foremost because the OPIS model was also designed and validated to predict cumulative live births over a number of IVF cycles. Additionally, it is based on more recent HFEA data and is therefore likely to capture possible improvements in effectiveness since IVF Predict was developed. Out of a concern for patient confidentiality, various age categories (18-34, 35-37, 38-39, 40-42) were used for the age predictor in IVF Predict but this leads to some anomalies when used to estimate the treatment effect relative to EM, where age is a continuous variable. Thus, while predicted live birth falls constantly in the EM prediction model, it is constant within a particular age range in IVF Predict. That means that within that IVF Predict age band, IVF appears more effective with increasing age, as a constant IVF live birth prediction is contrasted with a falling EM prediction. Finally, IVF Predict as used to estimate cumulative live birth for the economic model, could produce a counter intuitive result where the probability of live birth in subsequent cycles was higher than in previous cycles. Whilst this arises from using IVF Predict to estimate cumulative live birth over a number of cycles, for which it is not designed, it is nevertheless an important limitation, as worsening average prognosis and decline in the probability of live birth with increasing cycles would be expected, a phenomenon referred to as depletion of susceptibles.

Below, we describe in more detail the prediction models, including their predictors, equations and any adaptations made for the purposes of this economic analysis.

i. van Eekelen (base case)

The predictors in the van Eekelen model are as follows:

- Woman's age
- Duration of subfertility
- Primary subfertility (1 = yes; 0 = no)
- Percentage motile sperm
- Referred by a gynaecologist (1 = yes; 0 = no)

The formula used to estimate spontaneous conception leading to live birth is given by:

μ = Exp (PI)/(1+Exp (PI))

Where:

μ is the probability of spontaneous conception leading to live birth in the first menstrual cycle

PI is the Prognostic Index which is calculated as follows:

$$PI = constant + (\alpha_1 \times age1) + (\alpha_2 \times age1) + (\alpha_3 \times duration) + (\alpha_4 \times primary\ subfertility) + (\alpha_5 \times \%\ motile\ sperm) + (\alpha_6 \times Gynaecological\ referral)$$

Where:

age1 = age if age < 31 years *or*
age1 = 31 if age ≥ 31 years

age2 = 0 if age < 31 years *or*
age2 = 31 – age if age ≥ 31 years

The coefficients for the formula used to calculate PI are given in Table 8.

Table 8: PI formula coefficient values for van Eekelen prediction model

Coefficient	Value
Constant	-1.56
α_1	-0.037
α_2	-0.074
α_3	-0.241
α_4	-0.359
α_5	0.0062
α_6	-0.654

The following formula is then used to calculate subsequent probability of spontaneous conception leading to live birth after previously unsuccessful cycles.

$$P = \prod_{i=m}^{i=m+j-1} \frac{1-\mu+i \times 0.1}{1+i \times 0.1}$$

Where:

P = predicted cumulative probability over the next j cycles

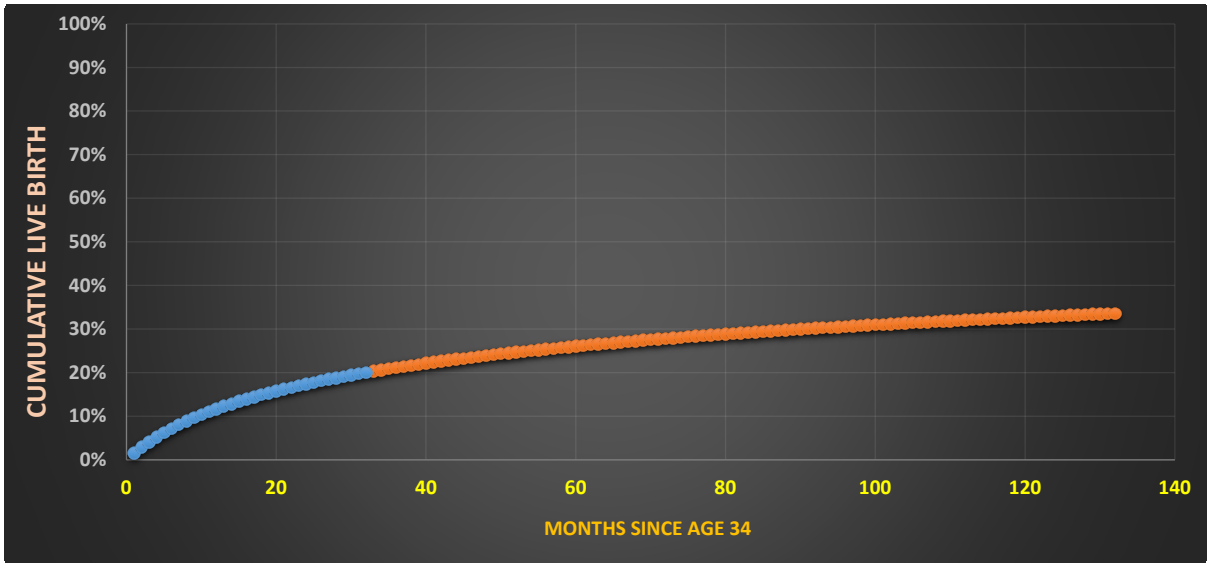
m = number of previous unsuccessful EM cycles

The authors report that this model was developed and internally validated over 32 cycles and therefore they recommend that predictions are not made for longer periods. However, given the time horizon of our model, most scenarios required prediction over a longer period and therefore this model was used to make predictions for the remainder of the women's reproductive life. The impact of this is illustrated in Figure 16 for a woman at a starting age of 34 years, with 2 years of trying to conceive, never having had a pregnancy, with a partner with 40% motile sperm and a gynaecology referral. The plot marked in blue represents the first 32 menstrual cycles with the orange representing predictions beyond that. Whilst recognising that there is greater uncertainty for predictions made beyond 32 months the logarithmic shape of the curve is in line with expectations about declining increases in cumulative live birth over time and therefore, we consider that these predictions, in the absence of any alternative, would still be useful for our purposes.

In addition to estimating the cumulative live birth rates for EM, the van Eekelen model was also used to predict spontaneous conception leading to live birth in women who had

completed their IVF treatment without achieving a live birth. In this the case the starting parameters for the van Eekelen prediction model would be adjusted to take account of a slightly older age and duration trying to conceive.

Figure 16: Illustration of predicted cumulative live birth rates from the van Eekelen model



ii. OPIS pre-IVF (base case)

The original OPIS pre-IVF model (McLernon 2016) was based on 113,873 women in the HFEA registry who started treatment between 1999 and 2008. The model was subsequently externally validated using more recent HFEA data in a population of 91,035 women who had IVF treatment between January 2010 and December 2016 (Ratna 2023). As a result, the model was recalibrated and the equation updated. We used the OPIS pre-IVF calculator based on these revised equations.

The predictors in the OPIS pre-IVF model are listed below:

- Woman's age
- Duration of subfertility
- Previous pregnancy (0 = yes; 1 = no)
- Tubal problem (1 = yes; 0 = no)
- Anovulation (1 = yes; 0 = no)
- Male factor (1 = yes, 0 = no)
- Unexplained (1 = yes; 0 = no)
- Treatment (1 = ICSI; 0 = IVF)

The formulas used in the OPIS pre-IVF model are outlined below:

Age1 = MAX((Age-26)/k_age,0)^3+(11*(MAX((Age-41)/k_age,0))^3-(15)*MAX((Age-37)/k_age,0)^3)/4

Age2 = MAX((Age -31)/k_age,0)^3+(6*MAX((Age -41)/k_age,0)^3-(10)*MAX((Age -37)/k_age,0)^3)/4

Age3 = MAX((Age-34)/k_age,0)^3+(3*MAX((Age -41)/k_age,0)^3-(7)*MAX((Age - 37)/k_age,0)^3)/4

Year1 = MAX((Year+9)/k_year,0)^3+((6)*MAX((Year)/k_year,0)^3-(9)*MAX((Year+3)/k_year,0)^3)/(3)

Year2 = MAX((Year+6)/k_year,0)^3+((3)*MAX((Year)/k_year,0)^3-(6)*MAX((Year+3)/k_year,0)^3)/(3)

Where:

Year = 0
K_age = 15^(2/3)
K_year = 9^(2/3)

Then calculate:

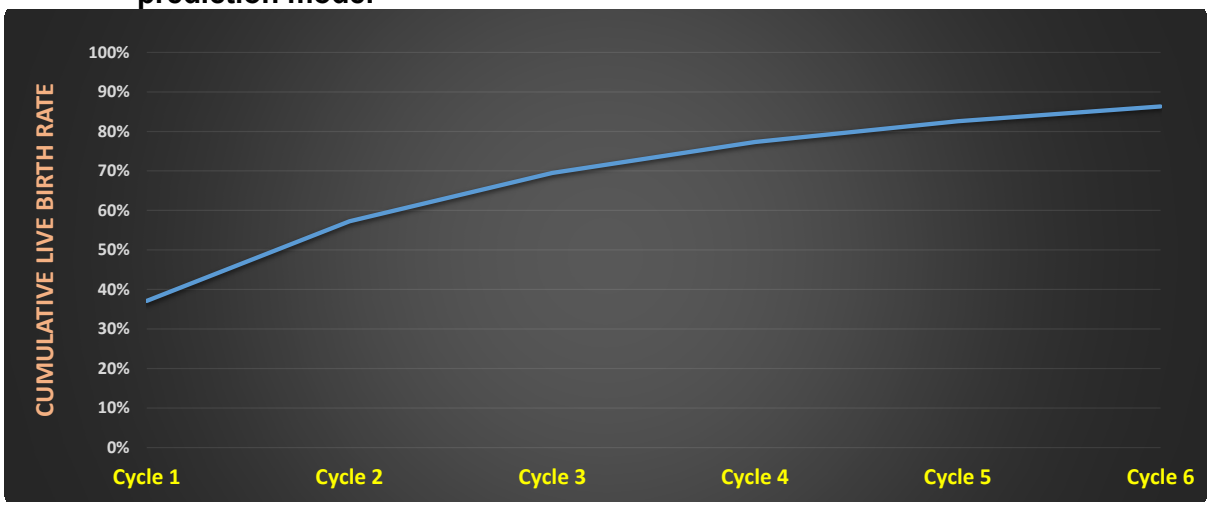
XB = -1.775 + (0.025*Age) + (-0.222*Age1_) + (0.732*Age2_) + (-1.804*Age3_) + (-0.016*Duration) + (-0.143*Previous_Preg) + (-0.091*Tubal) + (0.139*Anovulation) + (0.051*Male_factor) + (0.057*Unexplained) + (-0.006*ICSI) + (-0.111*Year) + (0.255*Year1) + (-0.587 *Year2)

The to calculate the probability of birth in each cycle:

P_{cycle1} = Exp (XB) / (1+Exp (XB))
P_{cycle2} = Exp (XB-0.226) / (1+Exp (XB-0.226))
P_{cycle3} = Exp (XB-0.388) / (1+Exp (XB-0.388))
P_{cycle4} = Exp (XB-0.531) / (1+Exp (XB-0.531))
P_{cycle5} = Exp (XB-0.679) / (1+Exp (XB-0.679))
P_{cycle6} = Exp (XB-0.768) / (1+Exp (XB-0.768))

An example of cumulative live birth rates across 6 cycles of IVF is depicted in Figure 17. It is for a woman starting treatment at age 34 years, with 2 years of trying to conceive, never having had a pregnancy, unknown cause, no male factor and not using ICSI.

Figure 17: Illustration of predicted cumulative live birth using the OPIS pre-IVF prediction model



iii. Hunault (sensitivity analysis)

The Hunault model was developed based on primary data from 2459 sub-fertile couples from 3 different studies (Eimers et al., 1994, Collins et al., 1995; Snick et al., 1997) and predicts spontaneous conception leading to live birth based on:

- Woman's age
- Duration of subfertility
- Primary subfertility (1 = yes; 0 = no)
- Percentage motile sperm
- Tertiary couple (1 = yes; 0 = no)

The formula used in this model to predict a spontaneous conception leading to live birth is given by:

$$P = (1 - 0.181^{\exp(PI)})$$

Where:

P is the predicted probability of spontaneous conception leading to live birth within 1 year

PI is the Prognostic Index, given by*:*

$PI = (\beta_1 \times AGE1) + (\beta_2 \times AGE2) + (\beta_3 \times \text{duration of subfertility}) +$
$(\beta_4 \times \text{primary subfertility}) + (\beta_5 \times \text{percentage of motile sperm}) +$
$(\beta_6 \times \text{tertiary-care couple})$

Where:

AGE1 is the woman's age if she is 31 years or younger, or 31 if the woman's age is more than 31 years.

AGE2 is the difference between woman's age and 31 years if the woman's age is more than 31 years and zero if the woman is 31 years or younger.

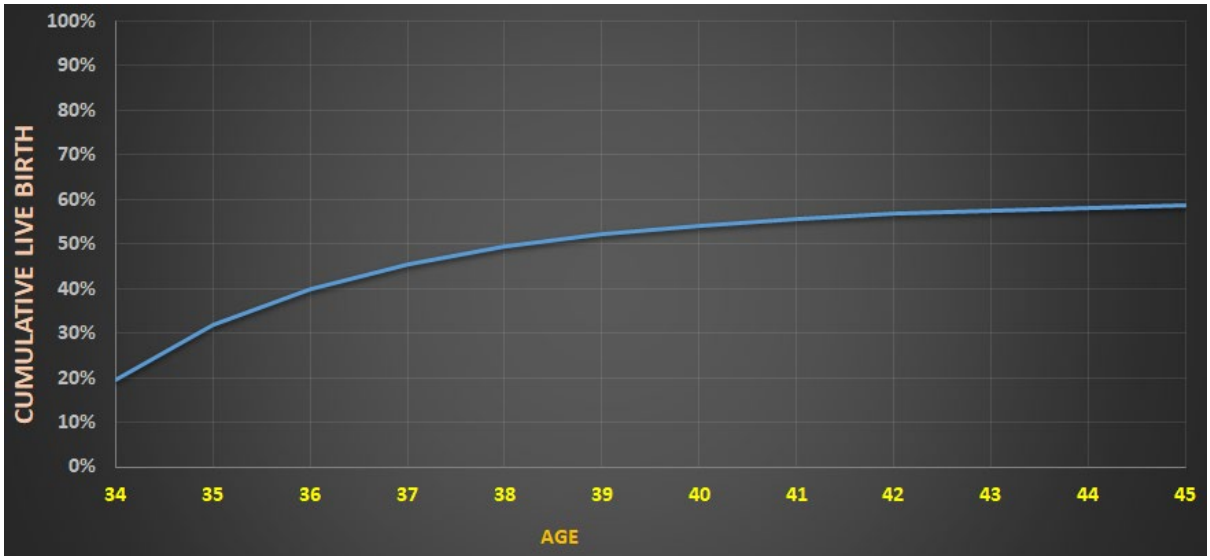
The coefficients for the formula used to calculate PI are given in Table 9.

Table 9: PI formula coefficient values for Hunault prediction model

Coefficient	Value
β_1	-0.03
β_2	-0.08
β_3	-0.19
β_4	-0.58
β_5	-0.008
β_6	-0.25

Cumulative live birth rates are estimated by incrementing the age and duration parameters for subsequent 12-month periods. Figure 18 gives an example of cumulative live birth rates for a woman at a starting age of 34 years, with 2 years of trying to conceive, never having had a pregnancy, with a partner with 40% motile sperm and in tertiary care.

Figure 18: Illustration of predicted cumulative live birth rates using the Hunault model



iv. IVF Predict (sensitivity analysis)

The IVF Predict model was based on data of 144,018 fresh IVF cycles undertaken in the UK between 2003 and 2007 held on the HFEA database. A multivariable logistic regression model was used to assess associations between pre-defined characteristics and live birth formed the basis of the prediction model.

Live birth is predicted from the following variables:

- Woman's age
- Duration of subfertility
- Own or donor eggs
- Cause (unknown, anovulatory only, endometriosis only, cervical only, low sperm count, combined causes)
- IVF attempts
- Unsuccessful IVF attempts
- Pregnancy history
- Medication

- ICSI used

The formula used in IVF Predict is as follows:

$P = \exp(y) \div (1+\exp[y])$

Where:

P is the probability of live birth

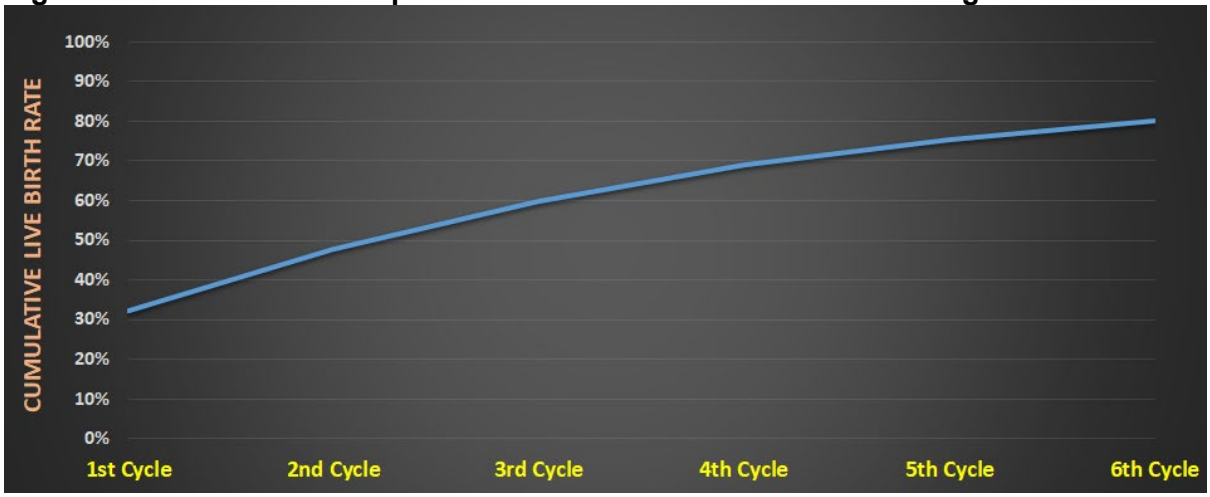
y = -1.1774 + (age and duration effect) + (age and source of embryo effect) + (ICSI and cause effect) + (ICSI and cycle number effect) + (previous number of unsuccessful IVF attempts) + (previous obstetric history effect) + (hormonal preparation effect)

The values for these effects can be found in tables produced as supplementary material (Text S2) to the published paper (Nelson 2011) which is available for download from: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000386#s6.

In order to derive cumulative live birth rates for subsequent cycles of IVF, age and duration parameters are incremented by the months since the previous cycle. Also, the parameters relating to IVF attempts and unsuccessful IVF attempts are also incremented for subsequent cycles. The outputs of IVF Predict do not always show a subsequent IVF attempt to have a lower probability of success than a previous attempt. In the absence of better patient selection with increasing cycles it would be expected that the pool of remaining infertile women would have a worsening average prognosis as the number of failed cycles increases. Therefore, in our analysis the probability of live birth in a cycle is constrained to not exceed the probability in a previous cycle.

Figure 19 gives an example of cumulative live birth rates across 6 IVF cycles for a woman starting treatment at age 34 years, with 2 years of trying to conceive, never having had a pregnancy, unknown cause, using her own eggs, using gonadotropin and not using ICSI.

Figure 19: Illustration of predicted cumulative live birth rates using the IVF Predict



Matching the model predictors

The effectiveness of IVF compared to treatment alternatives was estimated using the prediction models outlined above. The variables included for predicting the success of the intervention and the success of expectant management were not identical. Therefore, the inputs were set to make the populations for the different models as closely matched as possible but additional assumptions were introduced in this process:

- van Eekelen and Hunault output was based on being 'referred by a gynaecologist' rather than a GP as the population covered by the guideline would typically be under specialist care rather than primary care and would best match the populations in the IVF prediction model who would be in tertiary care.
- Model output assumed that it was the woman's own eggs rather than donor eggs. In IVF Predict this was a predictor.
- In the OPIS pre-IVF tool, prediction was based on a population using their own eggs. However, within IVF Predict the source of embryo (own or donor) could be varied. The model developed for this guideline was based on a population using their own eggs.
- IVF Predict also included the medication used for ovarian stimulation as a predictor. It was assumed for this analysis that the medication used for ovulation induction would be gonadotrophins as that reflects most UK practice.
- The van Eekelen, Hunault and OPIS pre-IVF models included primary subfertility as a dichotomous prediction variable. IVF Predict allowed pregnancy history to be classed in 6 different ways but to match IVF Predict with the other prediction models and the majority of the population covered by our guideline, we used "no previous IVF, no previous pregnancy".
- The IVF prediction models included a male factor cause (unspecified in OPIS pre-IVF and low sperm count in IVF Predict) as a dichotomous predictor whereas the spontaneous conception models included percentage motile sperm as a predictor. For male factor cause in the spontaneous conception model, we assumed that 40% sperm motility or higher excluded a male factor cause. This was based on the World Health Organization (WHO) reference characteristics for human semen (Cooper 2009). Where a diagnosis of low sperm count was used in IVF Predict or simple male factor cause in OPIS pre-IVF, a sperm motility of 20% was used in the van Eekelen and Hunault models. The simplifying assumption was made that sub-optimal sperm motility (more precisely a sub-optimal value of 20%) is likely to be associated with a low sperm count in terms of its predicted impact on live birth rates.
- The populations of the studies used to inform predictions of spontaneous conception comprised of couples with unexplained fertility. For analyses assessing a mild endometriosis cause, it was assumed that the spontaneous conception leading to live birth would be the same as for unexplained fertility.
- Tubal disease, anovulation, severe endometriosis and cervical causes were assumed to have a zero natural conception rate.

A summary of the model predictors is given in Table 10 and Figure 20 illustrates an example of the differences in cumulative live birth rates produced by the different prediction models for the following scenario:

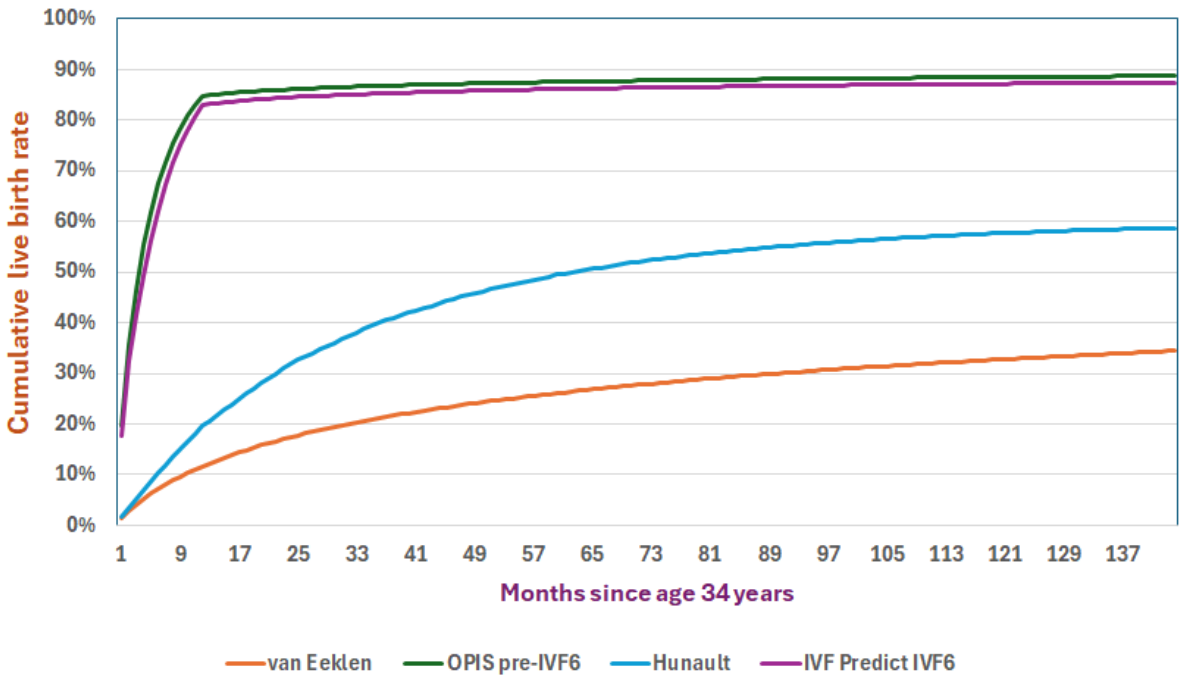
Woman age:	34 years
Duration trying to conceive:	2 years
Cause:	Unexplained
ICSI used:	No

IVF Cycles: 6

Table 10: Summary of model prediction variables

	Values			
Prediction variable	**van Eekelen**	**Hunault**	**OPIS pre-IVF**	**IVF Predict**
Age (years)	20-45	20-45	20-45	20-45
Duration (years)	2	2	2	2
Primary subfertility	1	1	1	1
Own/donor eggs	N/A	N/A	N/A	Own eggs
Referred by gynaecologist	Referred by gynaecologist	Referred by gynaecologist	N/A	N/A
Medication	N/A	N/A	N/A	Gonadotrophins
ICSI used	N/A	N/A	0 or 1	0 or 1
Previous IVF	N/A	N/A	N/A	0
Unexplained cause	N/A	N/A	0 or 1	0 or 1
Tubal cause	N/A	N/A	0 or 1	0 or 1
Anovulation cause	N/A	N/A	0 or 1	0 or 1
% motile sperm/male factor	20% or 40%	20% or 40%	0 or 1	0 or 1
Endometriosis cause	N/A	N/A	N/A	0 or 1
Cervical cause	N/A	N/A	N/A	0 or 1
Combined cause	N/A	N/A	N/A	0 or 1

Figure 20: Comparison of cumulative live birth rates for a woman aged 34 years and with unknown cause of subfertility using different prediction models



Note: No accounting for depletion of susceptibles and van Eekelen used to predict cumulative live births after unsuccessful completion of IVF

Depletion of susceptibles

The EM models were developed to predict live birth rates in a population who have not received IVF treatment. Whilst there may be a random or chance element to achieving a live birth in any given IVF cycle it will also be affected by prognosis. Women with a better prognosis within that cohort are more likely to have a baby. As a result, the average prognosis declines with each successive IVF treatment cycle which should be reflected in lower probabilities of live birth. This effect is known as the 'depletion of susceptibles' where those with the most favourable prognosis drop-out of the population.

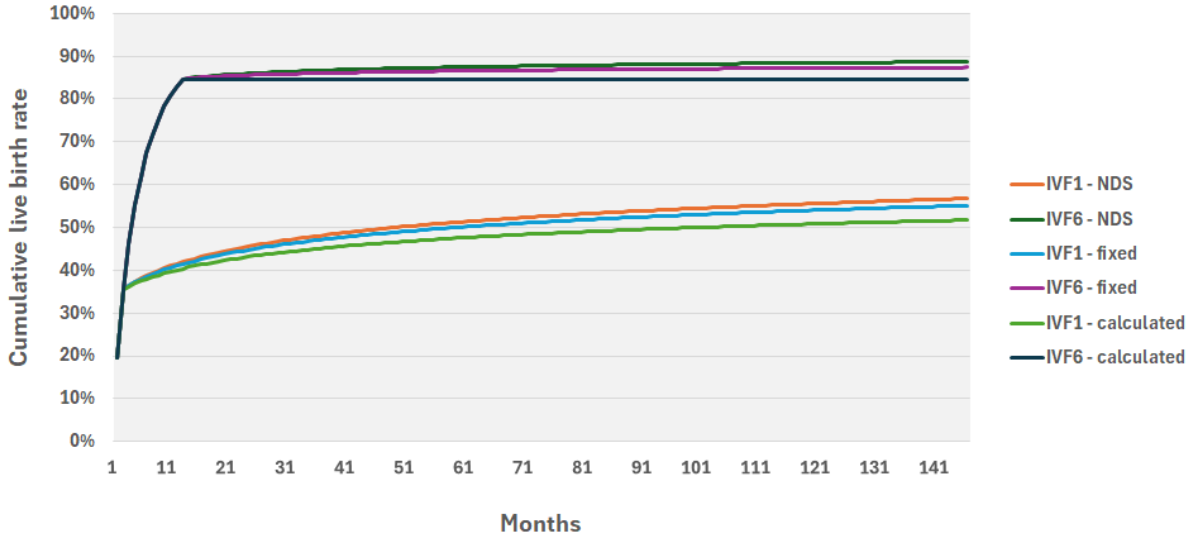
This is important for our model as we use the EM models to estimate cumulative live birth rates in women who have been treated with IVF and not had a live birth. Therefore, without adjustment the EM models are likely to over-estimate the chances of a spontaneous conception leading to a live birth in this population who are not treatment naïve as their unsuccessful IVF means they have a worse prognosis than the population used to inform the EM model predictions.

To take account of depletion of susceptibles it was assumed that a percentage of the remaining population who did not achieve a live birth with IVF have a zero probability of success with expectant management. We assumed that in our population there are 2 groups, one who can have a live birth and a group that cannot. In the group who can have a live birth the simplifying assumption was made that expectant management would be as effective as it would be in the treatment naive. Two approaches were made to estimating the proportion of the population who could not have a spontaneous conception following IVF treatment.

i. A fixed percentage, which was set to 5% in the base case analysis
ii. Using the proportion of the population estimated to not achieve a live birth after 6 cycles. This has the advantage that the proportion who cannot have a live birth increases with age as would be expected. However, it does mean that no additional live births occur in the group of women who do not have a live birth after 6 cycles of IVF.

Figure 21 is used to illustrate the impact of the different assumptions with respect to the depletion of susceptibles. It uses the example of a woman aged 34 years of age who has been trying to conceive for 2 years and whose cause of subfertility is unknown. The comparison is made for a strategy of 1 IVF cycle and 6 IVF cycles.

Figure 21: Impact on cumulative live birth rates of different depletion of susceptible assumption for a woman aged 34 years, trying to conceive for 2 years prior to starting IVF treatment and with subfertility of unknown cause using the OPIS pre-IVF and van Eekelen prediction models



IVF1 = 1 cycle of IVF; IVF6 = 6 cycles of IVF; NDS = no depletion of susceptibles; fixed = 5% of population assumed to have a zero probability of spontaneous conception leading to live birth; calculated = proportion assumed to have a zero probability of spontaneous conception calculated from proportion predicted not to achieve a live birth after 6 cycles of IVF

Costs and resource use

In line with the NICE guidelines manual, a NHS and personnel social services perspective was adopted for this analysis (https://www.nice.org.uk/process/pmg20). Costs were mostly based on a 2022/23 price year. However, NHS Reference Costs were based on the most recently available at the time of writing. In line with the NICE reference case, costs were discounted at a rate of 3.5% per annum.

Unit cost data that was only available from earlier years was updated for inflation to 2022/23 prices using the NHSCII (NHS Cost Inflation Index) and the now discontinued HCHS (Hospital and Community Health Service) index for pre-1917 costs (Unit Costs of Health and Social Care 2023).

Intervention costs are shown in Table 11. In costing IVF treatment, the model recognised that a proportion of cycles are cancelled and therefore may incur less costs. 2018-19 data from a personal communication with the HFEA was used to estimate the proportion of treatment cycles to attribute lower costs to arising from cancellation. These data are summarised in Table 12.

Table 11: Model intervention costs

Intervention	Cost	Source
Expectant management	£0.00	Guideline committee
IVF	£3,649	2023-25 NHS Payment Scheme (amended) [a] (https://www.england.nhs.uk/publication/2023-25-nhs-payment-scheme/#heading-2)
IVF with ICSI	£4,120	2023-25 NHS Payment Scheme (amended)

Intervention	Cost	Source
		(https://www.england.nhs.uk/publication/2023-25-nhs-payment-scheme/#heading-2)
IVF cancelled before egg collection	£1,224	NICE 2013 (CG156) and PSSRU Unit Costs Manual 2023 (https://www.pssru.ac.uk/unitcostsreport/#tabs_desc_10_2) [b]
IVF cancelled after egg collection	£3,140	Maheshwari 2010, NICE 2013 (CG156) and PSSRU Unit Costs Manual 2023 (https://www.pssru.ac.uk/unitcostsreport/#tabs_desc_10_2) [c]

(a) Price to include 1 fresh and 1 frozen cycle
(b) Cost from CG156 updated to 2022-23 prices using the NHSCII and HCHS inflation indices
(c) Cost from CG156 updated to 2022-23 prices using the NHSCII and HCHS inflation indices

Table 12: Proportion of cycles cancelled by treatment stage, 2018-19

Participant age (years)	Cycles cancelled at or before egg collection	Fresh embryo transfer cycles cancelled following egg collection	Cycles cancelled following egg/embryo thaw
18-36	4.1%	6.0%	1.7%

(a) Note to this data supplied by the HFEA: This data includes only NHS-funded IVF cycles begun with the intention of creating a pregnancy in 2018/19. Cycles cancelled at or before egg collection refers to the proportion of egg collections planned, and either abandoned or completed without collecting eggs. Fresh embryo transfer cycles cancelled following egg collection refers to the proportion of cycles with eggs collected but no subsequent embryos transferred. Cycles in which embryos were created, some were stored and/or donated, none were transferred, and some may have been discarded are excluded from this field. Cycles cancelled following egg/embryo thaw refers to the proportion of cycles in which an egg or embryo was thawed, but no embryo subsequently transferred. This data is from a live register and may not match data provided in previous requests or published elsewhere. One treatment centre has been excluded due to data quality issues.

The model also accounted for "downstream" costs related to birth, multiples in particular, and complications arising from OHSS. Differences in live birth rates between the intervention strategies will have implication for incremental costs. As fertility treatment is associated with higher multiple birth rates it was important to capture the additional costs that would be incurred by the NHS from the different strategies. However, as the NHS routinely provides treatment and care for births that were spontaneously conceived, a decision was made not to include the cost of singleton births in the base case analysis. Therefore, the cost of a multiple birth in the base case analysis was the incremental cost of a twin birth (for which there is more cost data than higher order multiples) compared to a singleton birth. So that this incremental cost can be estimated the costs of both singleton, and twin births are included as input parameters. The model is constructed so that singleton costs can be included as part of a sensitivity analysis in which case the multiple birth cost would be based on the absolute cost value for this input. The model inputs related to birth and OHSS costs are reported in Table 13.

Table 13: Birth and OHSS costs

Outcome	Cost	Source
OHSS moderate	£1,274	Maheshwari 2010, NICE 2013 (CG156) and PSSRU Unit Costs Manual 2023 (https://www.pssru.ac.uk/unitcostsreport/#tabs_desc_10_2) [a]
OHSS severe	£3,873	Maheshwari 2010, NICE 2013 (CG156) and PSSRU Unit Costs Manual 2023 (https://www.pssru.ac.uk/unitcostsreport/#tabs_desc_10_2) [a]
Singleton birth	£5,903	Jacklin and Marceniuk (2018) [b]
Multiple birth	£16,846	Jacklin and Marceniuk (2018) [b]

(a) Cost from CG156 updated to 2022-23 prices using the NHSCII and HCHS inflation indices

(b) https://www.hfea.gov.uk/media/2712/nga-twin-pregnancy-costing-final.pdf and updated to prices using the NHSCII

Health state utilities and QALYs

In accordance with the NICE reference case, an annual discount rate of 3.5% was applied to health state utilities accrued. Whilst the relationship between health-related quality of life and infertility is complex there is evidence to suggest that there are short term effects on mental health, depression and anxiety in particular, and longer-term effects on more general measures of well-being (Skedgel 2023).

Deriving health state utilities for live birth is not straightforward but for this analysis we have followed the approach in previous NICE guideline CG156 where any health state utility gain had to relate to a couple or person seeking treatment and not any, as yet unconceived life. In the base case analysis this utility gain was limited to the person giving birth but possible impacts on partner health-related quality of life were explored in a sensitivity analysis. It was assumed that any gain in health-related quality of life resulting from a live birth would persist for the remainder of the woman's life. Although this assumption could potentially over-estimate the overall QALY gain from a live birth it will be mitigated by the heavy discounting of health state utilities in later years of life. The model uses the same health state utility gain for a live birth as was used in the previous guideline (NICE 2013).

A health state utility loss was also applied for occurrences of moderate and severe OHSS. As no published health state utility values was found for these health states, they were proxied by using values for moderate and severe pain respectively. The model assumed a duration of 2 weeks for OHSS symptoms although 7-10 days is more typical (RCOG 2016). The model makes the simplifying assumption that there are no health state utility implications related to multiple birth over and above live birth more generally. This was because the most serious adverse outcomes associated with multiple birth and prematurity are more frequently experienced by the neonate than the mother and again this would involve estimating QALYs for an individual that did not exist at the point at which treatment is commenced.

The health state utilities used in the model are summarised in Table 14.

Table 14: Model health state utilities

Outcome	Health state utility gain	Duration	Source
OHSS moderate	-0.084	14 days	Devlin 2016 [a]
OHSS severe	-0.276	14 days	Devlin 2016 [b]
Live birth	0.070	Lifelong	Scotland 2011

(a) Based on an EQ-5D-5L value set for England and using moderate pain as a proxy for moderate OHSS
(b) Based on an EQ-5D-5L value set for England and using severe pain as a proxy for severe OHSS

Results are illustrated for cost-effectiveness thresholds of £20,000 per QALY and £30,000 per QALY. The higher £30,000 per QALY threshold is greater than the £20,000 per QALY suggested as a benchmark criteria for assessing cost-effectiveness in "NICE: Our Principles" but the committee believed that the QALY did not fully capture other non-HRQoL benefits of fertility treatment and that huge uncertainty existed with respect to both the method and valuation of treatment benefits. Other authors have also suggested that QALY approaches do not fully capture the benefits of fertility treatment (Keller 2022, Skedgel 2023).

However, decision making was more nuanced than just using a single threshold as the committee had to distil a small number of pragmatic recommendations derived from multiple analyses addressing different scenarios, inputs and assumptions. Therefore, the recommendations made were intended to reflect the totality of the cost-effectiveness evidence. To this end cost-effectiveness at a more stringent £20,000 cost per QALY was an

important consideration when addressing options that would extend treatment beyond current NICE guidance. However, a more permissive £30,000 per QALY threshold was used as a benchmark for assessing the cost-effectiveness of existing NICE recommendations. The committee also took into account limitations in the modelling approach and the sensitivity of model conclusions to some highly uncertain inputs, which could lead to a systematic under or over estimation of cost-effectiveness.

Incremental analysis

To assess cost-effectiveness, strategies are ranked in order of cost from lowest to highest and incremental cost-effectiveness ratios (ICERs) are calculated relative to the next most cost-effective alternative were further cycles not to be available.

Sensitivity analysis

In a departure from the NICE reference case, probabilistic sensitivity analysis was not undertaken as the prediction models do not report uncertainty around their outcome. However, a number of deterministic sensitivity analyses were undertaken to explore the following:

i. Choice of prediction model
ii. Impact of assumptions with respect to depletion of susceptibles
iii. Spacing between IVF cycles (4 months)
iv. Use of ICSI (for male factor)
v. Cost of IVF treatment (£6,000)
vi. Health state utility gain assigned to a live birth (0.05 to 0.14)
vii. Discount rate (1.5%)
viii. Inclusion of singleton birth costs

For tubal, anovulatory, severe endometriosis and cervical causes of subfertility, it was assumed in the base case analysis that spontaneous conception leading to live birth was not possible. Threshold analyses were undertaken to assess how robust the results were to relaxing this assumption.

Results

A technical note at the end of this document explains the presentation and interpretation of the results.

Base case analyses

Unexplained cause

Figure 22 and Figure 23 summarise the base case results for women with unexplained fertility. They show that for women aged 37 years and younger 6 cycles of IVF are cost-effective at a cost-effectiveness threshold of £20,000 per QALY. For women aged 38-39 years 6 cycles are also cost-effective but some of the latter cycles are only cost-effective if a more generous £30,000 per QALY cost-effectiveness threshold is used. At 40 years of age, 3 cycles of IVF are cost-effective but only at a cost-effectiveness threshold of £30,000 per QALY is used. IVF is not shown to be cost-effective for women older than 41 years.

Figure 22: ICERs for base case analysis by age and IVF cycle for unexplained cause

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£14,938	£13,913	£15,007	£16,044	£17,297	£18,277
21	£14,077	£13,296	£14,341	£15,332	£16,527	£17,462
22	£13,301	£12,724	£13,723	£14,672	£15,814	£16,708
23	£12,601	£12,194	£13,150	£14,059	£15,152	£16,007
24	£11,966	£11,701	£12,617	£13,490	£14,537	£15,356
25	£11,388	£11,243	£12,121	£12,960	£13,965	£14,751
26	£10,861	£10,816	£11,659	£12,466	£13,431	£14,186
27	£10,385	£10,424	£11,236	£12,014	£12,943	£13,671
28	£9,986	£10,097	£10,885	£11,643	£12,546	£13,253
29	£9,690	£9,865	£10,643	£11,393	£12,286	£12,987
30	£9,522	£9,757	£10,543	£11,303	£12,210	£12,923
31	£9,548	£9,822	£10,614	£11,380	£12,297	£13,010
32	£9,510	£9,893	£10,741	£11,568	£12,560	£13,337
33	£9,613	£10,110	£11,032	£11,939	£13,030	£13,888
34	£9,774	£10,387	£11,391	£12,385	£13,580	£14,525
35	£9,955	£10,684	£11,770	£12,852	£14,154	£15,185
36	£10,351	£11,218	£12,425	£13,631	£15,085	£16,238
37	£11,299	£12,368	£13,796	£15,232	£16,964	£18,335
38	£13,301	£14,698	£16,553	£18,427	£20,694	£22,476
39	£16,920	£18,818	£21,426	£24,076	£27,291	£29,792
40	£23,052	£25,666	£29,552	£33,527	£35,680	£37,425
41	£33,133	£34,720	£36,804	£38,923	£41,117	£43,163
42	£49,351	£51,467	£54,719	£58,134	£61,743	£65,127
43	£75,893	£78,626	£83,841	£89,511	£95,636	£101,407
44	£120,900	£124,993	£134,225	£144,622	£156,199	£167,303
45	£205,522	£219,831	£244,943	£275,572	£314,548	£360,611

1

Figure 23: Cost-effectiveness of IVF by age and IVF cycles for unexplained cause

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

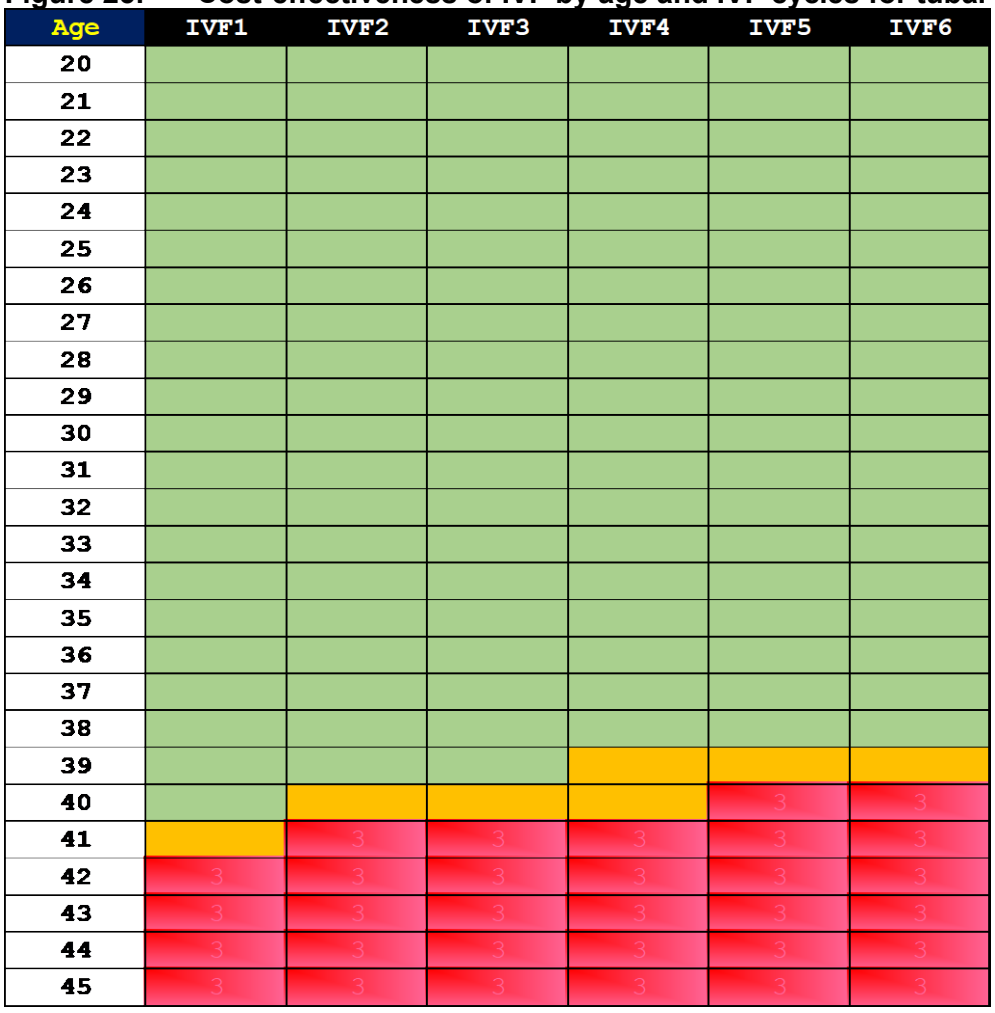
Tubal cause

It was assumed that spontaneous conception leading to live birth was not possible for tubal causes of infertility and therefore no EM prediction model was required for this analysis. The results for the base case analysis are shown in Figure 24 and Figure 25. Six cycles of IVF are cost-effective for women aged 38 years and younger. For women aged 39 years, 6 cycles are also cost-effective but for more than 3 cycles this is only the case with a cost-effectiveness threshold of £30,000 per QALY. At age 40, then 1 cycle is cost-effective at the more stringent £20,000 per QALY cost-effectiveness threshold but this increases to 4 cycles if the more generous threshold of £30,000 per QALY is used. A single cycle of IVF is cost-effective at age 41 years, but only at a cost-effectiveness threshold of £30,000 per QALY.

Figure 24: ICERs for base case analysis by age and IVF cycle for tubal cause

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£5,933	£6,875	£7,694	£8,535	£9,543	£10,386
21	£5,866	£6,789	£7,591	£8,416	£9,403	£10,230
22	£5,803	£6,707	£7,493	£8,301	£9,268	£10,078
23	£5,742	£6,628	£7,398	£8,190	£9,138	£9,932
24	£5,683	£6,552	£7,307	£8,083	£9,012	£9,791
25	£5,627	£6,479	£7,219	£7,979	£8,890	£9,654
26	£5,574	£6,409	£7,134	£7,880	£8,773	£9,523
27	£5,526	£6,346	£7,058	£7,790	£8,667	£9,404
28	£5,500	£6,309	£7,013	£7,736	£8,602	£9,330
29	£5,513	£6,322	£7,025	£7,748	£8,614	£9,342
30	£5,585	£6,408	£7,124	£7,860	£8,741	£9,482
31	£5,738	£6,597	£7,343	£8,110	£9,029	£9,801
32	£5,993	£6,913	£7,712	£8,533	£9,517	£10,342
33	£6,330	£7,331	£8,200	£9,094	£10,165	£11,060
34	£6,707	£7,799	£8,748	£9,724	£10,892	£11,867
35	£7,105	£8,293	£9,326	£10,387	£11,658	£12,717
36	£7,659	£8,983	£10,133	£11,316	£12,732	£13,909
37	£8,623	£10,186	£11,545	£12,941	£14,613	£15,999
38	£10,390	£12,395	£14,138	£15,930	£18,075	£19,846
39	£13,383	£16,141	£18,539	£21,004	£23,955	£26,382
40	£18,206	£22,182	£25,638	£29,191	£33,444	£35,041
41	£25,687	£31,552	£33,818	£35,977	£38,176	£40,210
42	£36,860	£40,610	£43,862	£46,899	£49,919	£52,753
43	£53,430	£59,040	£63,927	£68,513	£73,094	£77,409
44	£78,023	£86,395	£93,712	£100,601	£107,503	£114,017
45	£114,550	£127,026	£137,956	£148,267	£158,619	£168,405

Figure 25: Cost-effectiveness of IVF by age and IVF cycles for tubal cause



Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

2

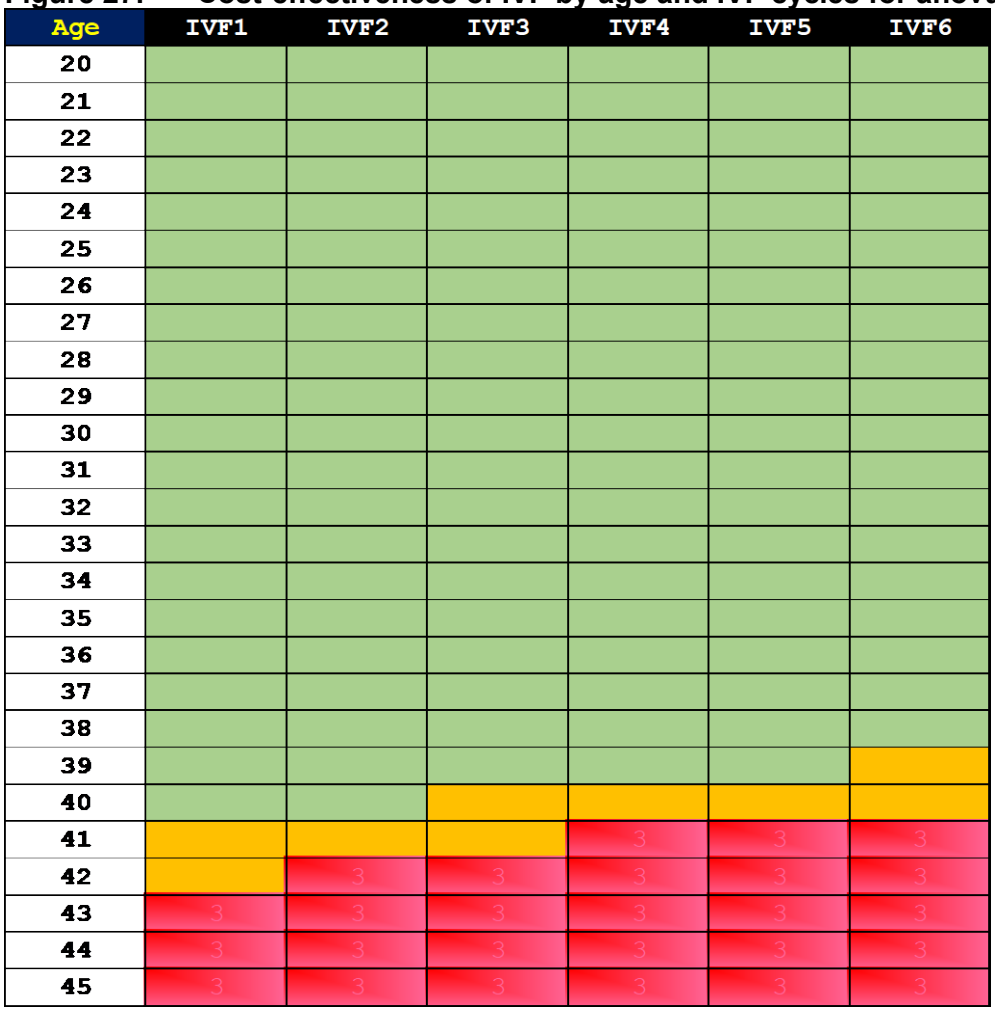
Anovulatory cause

It was assumed that spontaneous conception leading to live birth was not possible for anovulatory causes of infertility and therefore no EM prediction model was required for this analysis. The base case results for anovulatory cause are outlined in Figure 26 and Figure 27. For women aged 38 years and younger, 6 cycles of IVF are cost-effective at £20,000 per QALY. Between 39-42 years, some amount of IVF is cost-effective although the precise amount varies with age and the cost-effectiveness threshold. Whilst all 6 cycles are nearly cost-effective for women aged 39 at a £20,000 per QALY cost-effectiveness threshold, a single cycle is only borderline cost-effective for women aged 42 years when utilising a decision cost-effectiveness threshold of £30,000 per QALY.

Figure 26: ICERs for base case analysis by age and IVF cycle for anovulatory cause

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£5,169	£5,918	£6,568	£7,237	£8,038	£8,712
21	£5,119	£5,852	£6,489	£7,144	£7,929	£8,590
22	£5,070	£5,789	£6,413	£7,055	£7,824	£8,472
23	£5,024	£5,728	£6,340	£6,969	£7,722	£8,358
24	£4,980	£5,670	£6,269	£6,886	£7,624	£8,248
25	£4,938	£5,614	£6,202	£6,806	£7,530	£8,142
26	£4,898	£5,561	£6,137	£6,730	£7,439	£8,040
27	£4,862	£5,513	£6,079	£6,661	£7,358	£7,948
28	£4,844	£5,487	£6,046	£6,621	£7,309	£7,892
29	£4,857	£5,500	£6,059	£6,634	£7,321	£7,905
30	£4,917	£5,572	£6,140	£6,725	£7,425	£8,019
31	£5,042	£5,724	£6,318	£6,927	£7,657	£8,275
32	£5,248	£5,979	£6,614	£7,267	£8,048	£8,708
33	£5,519	£6,314	£7,005	£7,716	£8,566	£9,283
34	£5,823	£6,690	£7,444	£8,220	£9,147	£9,927
35	£6,143	£7,086	£7,907	£8,750	£9,760	£10,607
36	£6,587	£7,638	£8,552	£9,492	£10,617	£11,558
37	£7,357	£8,599	£9,678	£10,788	£12,116	£13,222
38	£8,765	£10,358	£11,743	£13,167	£14,871	£16,284
39	£11,148	£13,339	£15,245	£17,203	£19,548	£21,482
40	£14,985	£18,144	£20,890	£23,713	£27,092	£29,867
41	£20,934	£25,594	£29,646	£33,811	£36,047	£37,938
42	£29,817	£36,723	£39,403	£41,964	£44,580	£47,006
43	£42,988	£47,423	£51,277	£54,884	£58,478	£61,856
44	£62,534	£69,163	£74,947	£80,382	£85,819	£90,944
45	£91,562	£101,452	£110,106	£118,259	£126,436	£134,160

Figure 27: Cost-effectiveness of IVF by age and IVF cycles for anovulatory cause



Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

Male factor cause

Figure 28 and Figure 29 suggest that IVF is not cost-effective in women aged 41 years and above for male factor cause in the base case analysis. Six cycles of IVF are cost-effective at a cost-effectiveness threshold of £20,000 per QALY providing the woman is 37 years and younger. For women aged 38-39 years, six cycles are also cost-effective though some of the higher order cycles are only cost-effective using a £30,000 per QALY cost-effectiveness threshold. At age 40, three cycles can be considered cost-effective if using a more permissive cost-effectiveness threshold of £30,000 per QALY.

Figure 28: ICERs for base case analysis by age and IVF cycle for male factor cause

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£13,259	£12,778	£13,835	£14,847	£16,068	£17,028
21	£12,576	£12,254	£13,265	£14,235	£15,404	£16,323
22	£11,956	£11,767	£12,736	£13,667	£14,786	£15,667
23	£11,391	£11,314	£12,243	£13,138	£14,211	£15,057
24	£10,875	£10,891	£11,784	£12,644	£13,675	£14,488
25	£10,402	£10,497	£11,355	£12,184	£13,174	£13,956
26	£9,968	£10,129	£10,955	£11,753	£12,707	£13,460
27	£9,575	£9,791	£10,588	£11,359	£12,279	£13,006
28	£9,246	£9,511	£10,286	£11,037	£11,933	£12,641
29	£9,008	£9,318	£10,083	£10,827	£11,714	£12,416
30	£8,884	£9,241	£10,014	£10,768	£11,668	£12,381
31	£8,933	£9,324	£10,106	£10,869	£11,783	£12,500
32	£8,946	£9,432	£10,268	£11,089	£12,074	£12,853
33	£9,087	£9,676	£10,585	£11,484	£12,564	£13,421
34	£9,283	£9,978	£10,967	£11,950	£13,133	£14,074
35	£9,497	£10,300	£11,370	£12,440	£13,726	£14,752
36	£9,915	£10,850	£12,039	£13,231	£14,666	£15,811
37	£10,857	£11,996	£13,402	£14,819	£16,527	£17,886
38	£12,808	£14,287	£16,110	£17,955	£20,183	£21,946
39	£16,302	£18,318	£20,872	£23,471	£26,619	£29,085
40	£22,184	£24,992	£28,781	£32,659	£34,755	£36,462
41	£31,775	£33,507	£35,586	£37,668	£39,808	£41,803
42	£47,033	£49,440	£52,675	£56,005	£59,491	£62,759
43	£71,611	£74,931	£80,075	£85,528	£91,348	£96,828
44	£112,320	£117,469	£126,336	£136,019	£146,624	£156,763
45	£185,458	£199,743	£221,381	£246,682	£277,612	£312,658

Figure 29: Cost-effectiveness of IVF by age and IVF cycles for male factor cause

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

Mild endometriosis

As endometriosis was not a predictor in the OPIS pre-IVF tool, IVF Predict was used in this analysis. Broadly speaking this shows that for mild endometriosis, six cycles of IVF are cost-effective at a cost-effectiveness threshold of £20,000 per QALY for women aged 39 years and younger in the base case scenario. However, above this age no amount of IVF appears cost-effective (excepting the anomaly of 1 cycle of IVF with an ICER of £29,553 at age 42 years, which is explained in the Discussion). This is illustrated in Figure 30 and Figure 31.

Figure 30: ICERs for base case analysis by age and IVF cycle for mild endometriosis

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£18,763	£18,958	£18,097	£17,295	£18,754	£17,850
21	£17,958	£18,418	£17,612	£16,860	£18,301	£17,454
22	£17,225	£17,913	£17,158	£16,453	£17,877	£17,083
23	£16,555	£17,440	£16,733	£16,071	£17,480	£16,734
24	£15,942	£16,998	£16,335	£15,714	£17,107	£16,406
25	£15,380	£16,583	£15,962	£15,378	£16,757	£16,098
26	£14,863	£16,194	£15,611	£15,063	£16,429	£15,809
27	£14,386	£15,830	£15,282	£14,767	£16,120	£15,537
28	£13,945	£15,488	£14,974	£14,489	£15,830	£15,281
29	£13,537	£15,166	£14,683	£14,227	£15,558	£15,041
30	£13,158	£14,864	£14,410	£13,981	£15,301	£14,815
31	£12,890	£14,649	£14,170	£13,722	£14,998	£14,505
32	£12,268	£14,120	£13,694	£13,295	£14,554	£14,113
33	£11,728	£13,651	£13,272	£12,915	£14,159	£13,764
34	£11,257	£13,235	£12,896	£12,578	£13,808	£13,455
35	£16,214	£19,420	£18,928	£18,472	£20,651	£20,120
36	£15,566	£18,857	£18,419	£18,013	£20,152	£19,678
37	£14,997	£18,355	£17,966	£17,603	£19,708	£19,285
38	£14,153	£17,473	£17,136	£16,822	£18,831	£18,465
39	£13,723	£17,085	£16,785	£16,504	£18,488	£18,160
40	£32,270	£35,639	£36,732	£37,151	£38,212	£38,845
41	£30,839	£34,288	£35,454	£35,937	£37,004	£37,658
42	£29,553	£38,151	£37,834	£37,540	£38,507	£38,993
43	£96,547	£106,616	£109,521	£110,357	£114,204	£116,304
44	£87,520	£99,011	£102,859	£104,400	£108,509	£110,985
45	£209,841	£249,556	£265,458	£274,459	£296,882	£316,089

Figure 31: Cost-effectiveness of IVF by age and IVF cycles for mild endometriosis

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

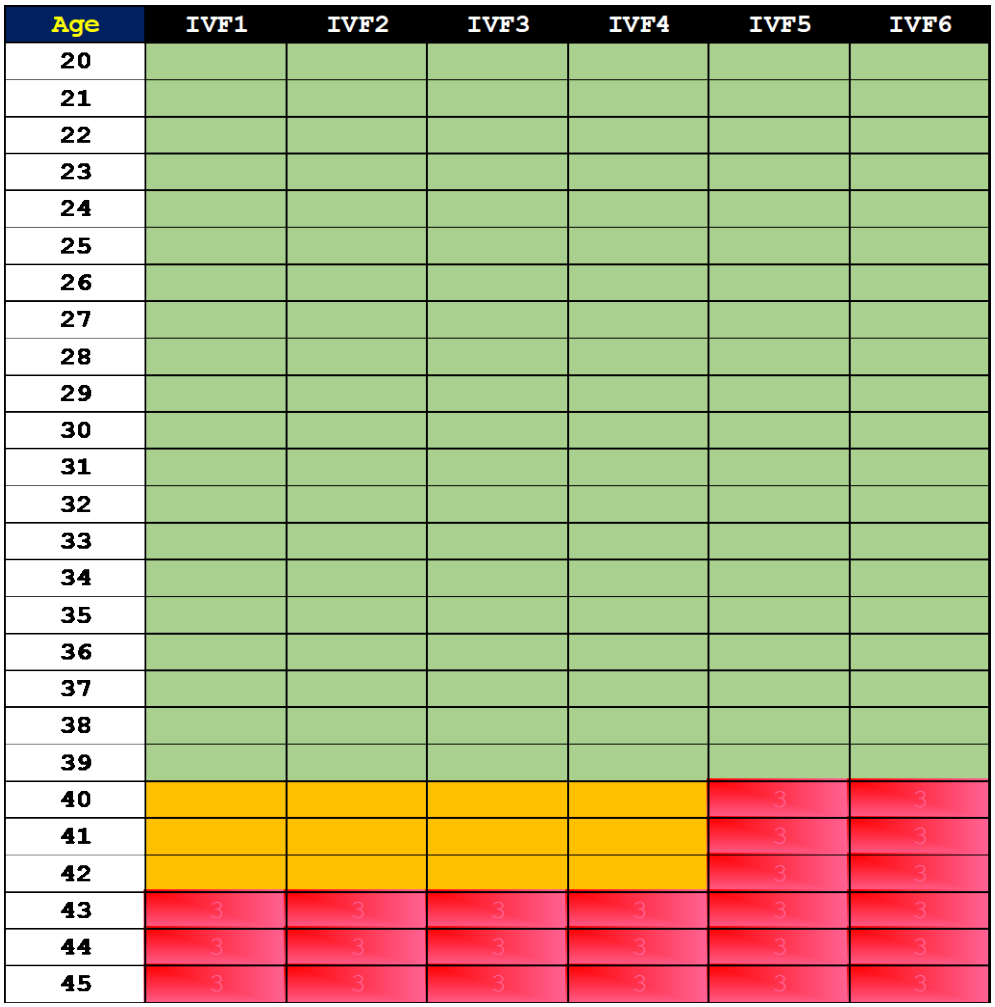
Severe endometriosis

As endometriosis was not a predictor in the OPIS pre-IVF tool, IVF Predict was used in this analysis. It was also assumed that spontaneous conception leading to live birth was not possible for severe endometriosis and therefore no EM prediction model was required for this analysis. As shown in Figure 32 and Figure 33, six cycles of IVF appear to be cost-effective in the base case analysis for women aged 39 years or younger with severe endometriosis at a cost-effectiveness threshold of £20,000 per QALY. For women aged 40-42 years then up to 4 cycles of IVF are cost-effective but only if a higher £30,000 per QALY cost-effectiveness threshold is deployed.

Figure 32: ICERs for base case analysis by age and IVF cycle for severe endometriosis

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£6,377	£8,146	£8,146	£8,146	£9,088	£9,088
21	£6,405	£8,181	£8,181	£8,181	£9,128	£9,128
22	£6,434	£8,218	£8,218	£8,218	£9,170	£9,170
23	£6,464	£8,257	£8,257	£8,257	£9,213	£9,213
24	£6,496	£8,298	£8,298	£8,298	£9,258	£9,258
25	£6,529	£8,340	£8,340	£8,340	£9,306	£9,306
26	£6,564	£8,385	£8,385	£8,385	£9,355	£9,355
27	£6,601	£8,431	£8,431	£8,431	£9,407	£9,407
28	£6,639	£8,480	£8,480	£8,480	£9,461	£9,461
29	£6,678	£8,531	£8,531	£8,531	£9,518	£9,518
30	£6,720	£8,584	£8,584	£8,584	£9,578	£9,578
31	£6,764	£8,640	£8,640	£8,640	£9,640	£9,640
32	£6,810	£8,699	£8,699	£8,699	£9,705	£9,705
33	£6,858	£8,760	£8,760	£8,760	£9,774	£9,774
34	£6,909	£8,825	£8,825	£8,825	£9,846	£9,846
35	£9,776	£12,904	£12,904	£12,904	£14,571	£14,571
36	£9,855	£13,007	£13,007	£13,007	£14,688	£14,688
37	£9,937	£13,116	£13,116	£13,116	£14,811	£14,811
38	£9,818	£12,937	£12,937	£12,937	£14,600	£14,600
39	£9,908	£13,056	£13,056	£13,056	£14,734	£14,734
40	£21,042	£28,914	£28,914	£28,914	£33,111	£33,111
41	£21,252	£29,202	£29,202	£29,202	£33,441	£33,441
42	£21,474	£29,507	£29,507	£29,507	£33,790	£33,790
43	£54,346	£63,258	£66,952	£68,972	£71,897	£73,998
44	£54,967	£63,981	£67,717	£69,760	£72,719	£74,843
45	£100,198	£117,134	£124,165	£128,015	£133,652	£137,704

Figure 33: Cost-effectiveness of IVF by age and IVF cycles for severe endometriosis



Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

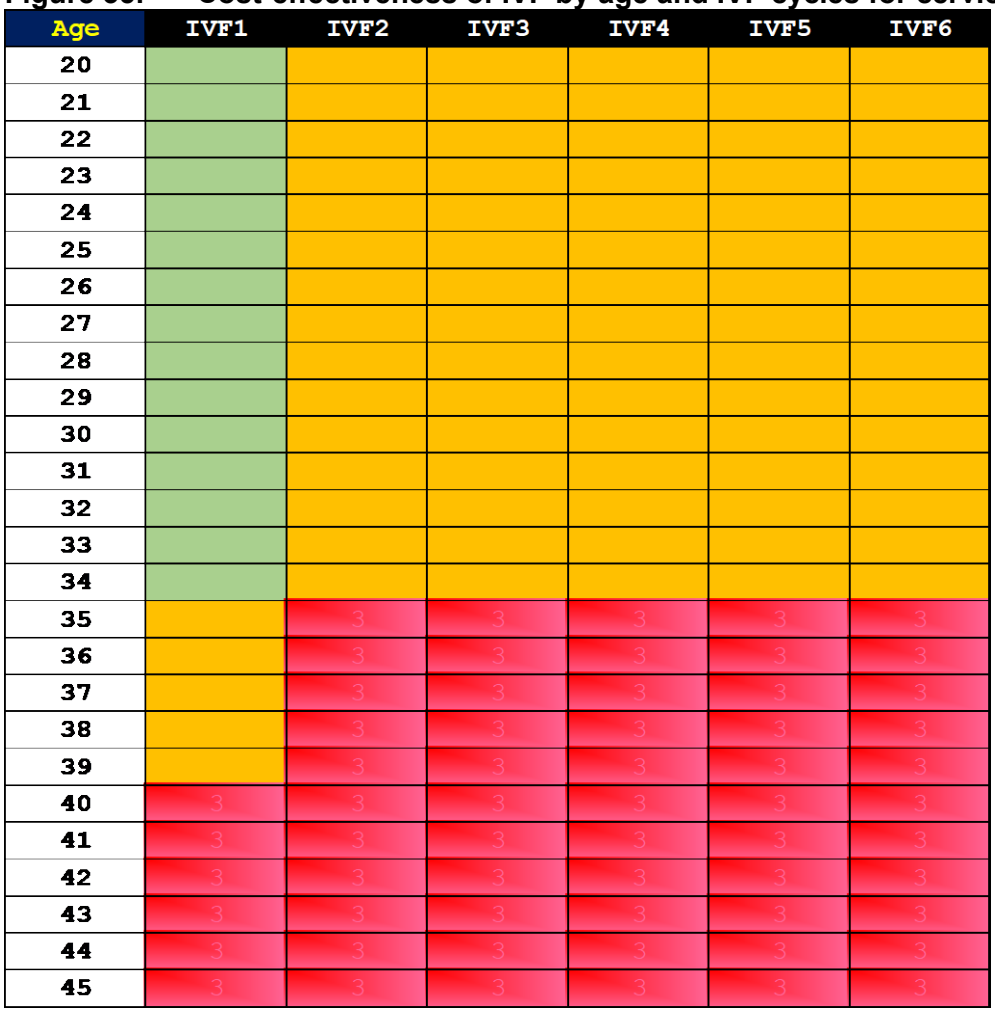
Cervical cause

As cervical cause was not a predictor in the OPIS pre-IVF tool, IVF Predict was used in this analysis. It was assumed that spontaneous conception leading to live birth was not possible for cervical causes of infertility and therefore no EM prediction model was required for this analysis. The base case results for cervical cause can be seen in Figure 34 and Figure 35. Under the ages of 35, then just 1 cycle of IVF is cost-effective at a cost-effectiveness threshold of £20,000 per QALY but 6 cycles can be considered cost-effective with a £30,000 per QALY cost-effectiveness threshold. For women aged 35-39 years a single cycle of IVF is cost-effective at a cost-effectiveness threshold of £30,000 per QALY.

Figure 34: ICERs for base case analysis by age and IVF cycle for cervical cause

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£14,883	£20,269	£20,269	£20,269	£23,140	£23,140
21	£14,948	£20,357	£20,357	£20,357	£23,241	£23,241
22	£15,016	£20,450	£20,450	£20,450	£23,347	£23,347
23	£15,087	£20,547	£20,547	£20,547	£23,457	£23,457
24	£15,162	£20,648	£20,648	£20,648	£23,572	£23,572
25	£15,239	£20,753	£20,753	£20,753	£23,693	£23,693
26	£15,320	£20,864	£20,864	£20,864	£23,819	£23,819
27	£15,405	£20,979	£20,979	£20,979	£23,951	£23,951
28	£15,494	£21,100	£21,100	£21,100	£24,089	£24,089
29	£15,587	£21,227	£21,227	£21,227	£24,234	£24,234
30	£15,684	£21,360	£21,360	£21,360	£24,385	£24,385
31	£15,787	£21,499	£21,499	£21,499	£24,544	£24,544
32	£15,894	£21,645	£21,645	£21,645	£24,711	£24,711
33	£16,006	£21,798	£21,798	£21,798	£24,886	£24,886
34	£16,124	£21,959	£21,959	£21,959	£25,069	£25,069
35	£24,819	£34,342	£34,342	£34,342	£35,375	£36,029
36	£25,018	£34,617	£34,617	£34,617	£35,659	£36,318
37	£25,227	£34,907	£34,907	£34,907	£35,958	£36,622
38	£24,821	£34,319	£34,319	£34,319	£35,346	£35,996
39	£25,047	£34,632	£34,632	£34,632	£35,669	£36,325
40	£58,903	£68,633	£72,667	£74,874	£78,078	£80,380
41	£59,490	£69,317	£73,392	£75,621	£78,857	£81,182
42	£60,111	£70,041	£74,158	£76,410	£79,680	£82,029
43	£160,148	£187,608	£199,019	£205,269	£214,469	£221,085
44	£161,978	£189,752	£201,294	£207,615	£216,921	£223,612
45	£299,642	£351,543	£373,124	£384,948	£402,416	£414,981

Figure 35: Cost-effectiveness of IVF by age and IVF cycles for cervical cause



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

2

Combined causes

As combined causes were not a predictor in the OPIS pre-IVF tool, IVF Predict was used in this analysis. Figure 36 and Figure 37 give the base case results for combined causes of subfertility. A combination of male factor and mild endometriosis causes were chosen for this example as neither of those causes utilised the assumption of zero probability of spontaneous conception leading to live birth. It suggests that IVF is not cost-effective for women aged 40 years and older. For women aged 34 years old and younger, then 6 cycles are cost-effective even at a cost-effectiveness threshold of £20,000 per QALY. Between ages 35-39, six cycles are also cost-effective, but for the higher order cycles then a £30,000 cost per QALY threshold would need to be used for decision making.

Figure 36: ICERs for base case analysis by age and IVF cycle for combined cause

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£19,186	£19,959	£19,137	£18,373	£20,140	£19,265
21	£18,434	£19,441	£18,670	£17,953	£19,695	£18,873
22	£17,748	£18,956	£18,233	£17,559	£19,277	£18,505
23	£17,120	£18,502	£17,824	£17,190	£18,886	£18,160
24	£16,544	£18,077	£17,440	£16,844	£18,520	£17,836
25	£16,014	£17,680	£17,080	£16,520	£18,176	£17,533
26	£15,525	£17,307	£16,743	£16,215	£17,854	£17,248
27	£15,074	£16,957	£16,427	£15,930	£17,551	£16,980
28	£14,656	£16,629	£16,130	£15,662	£17,268	£16,730
29	£14,269	£16,321	£15,852	£15,410	£17,001	£16,494
30	£13,909	£16,031	£15,590	£15,174	£16,751	£16,273
31	£13,669	£15,847	£15,378	£14,941	£16,474	£15,986
32	£13,069	£15,330	£14,912	£14,521	£16,031	£15,594
33	£12,547	£14,873	£14,499	£14,149	£15,638	£15,246
34	£12,091	£14,467	£14,132	£13,818	£15,290	£14,938
35	£17,804	£21,621	£21,119	£20,653	£23,234	£22,681
36	£17,148	£21,052	£20,604	£20,188	£22,723	£22,228
37	£16,574	£20,547	£20,148	£19,776	£22,270	£21,826
38	£15,672	£19,595	£19,249	£18,927	£21,307	£20,924
39	£15,240	£19,207	£18,898	£18,610	£20,961	£20,617
40	£36,924	£40,917	£42,233	£42,754	£44,048	£44,830
41	£35,353	£39,442	£40,844	£41,436	£42,736	£43,542
42	£33,947	£38,114	£39,590	£40,246	£41,552	£42,379
43	£114,366	£126,269	£129,681	£130,647	£135,321	£137,868
44	£103,561	£117,280	£121,881	£123,731	£128,741	£131,767
45	£256,328	£306,046	£326,145	£337,719	£367,080	£392,657

Figure 37: Cost-effectiveness of IVF by age and IVF cycles for combined cause

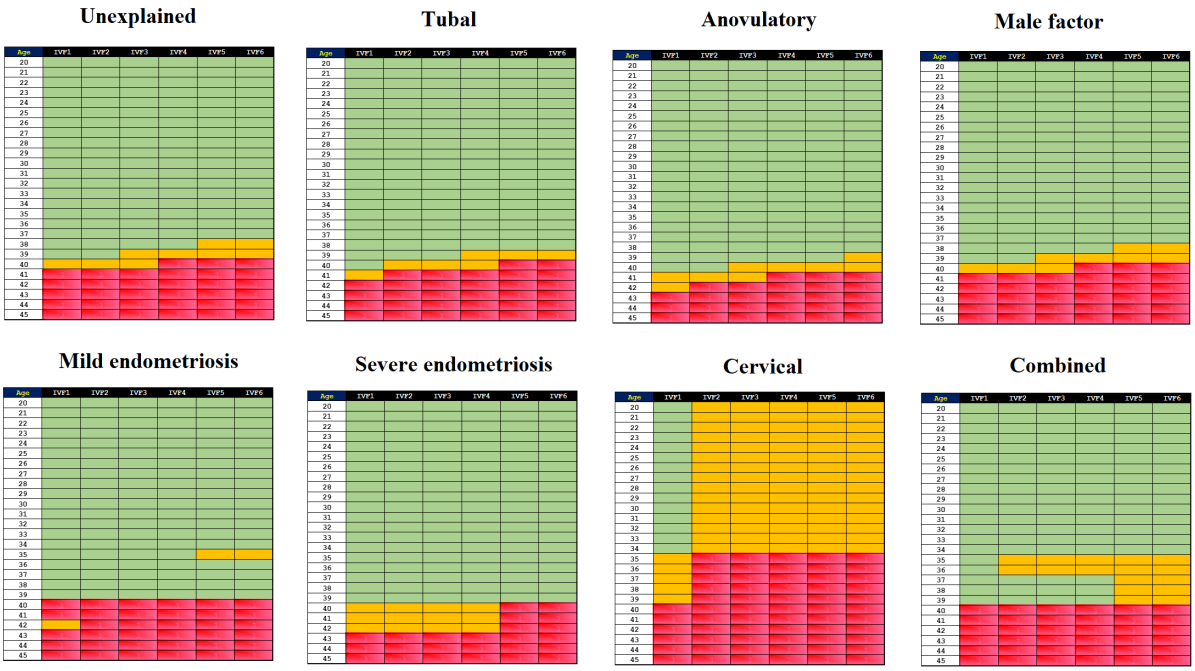
Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

Figure 38 summarises the cost-effectiveness for different causes of subfertility in the base case analyses.

Figure 38: Summary charts of cost-effectiveness of IVF cycles by age for the base case analysis using the OPIS pre-IVF (or IVF Predict) and van Eekelen prediction models



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Sensitivity analyses

Choice of model

In this sensitivity analysis results are presented for 6 base case scenarios using the IVF Predict and, where applicable, Hunault prediction models.

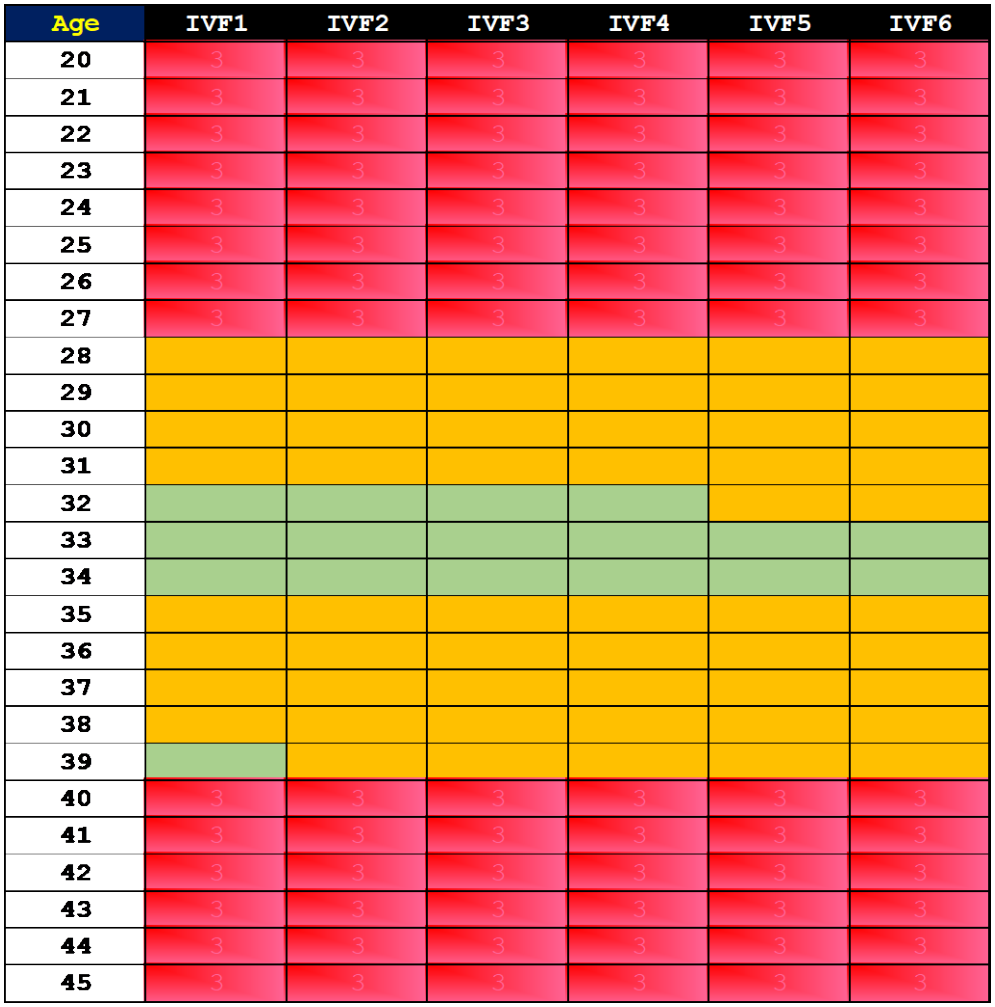
Unexplained cause

For this sensitivity analysis, the results for unexplained fertility are summarised in Figure 39 and Figure 40. IVF is not found to be cost-effective for women aged 27 years and younger or women aged 40 years or older. For women aged 28-39, six cycles are cost-effective, but this mostly requires that a more relaxed cost-effectiveness threshold of £30,000 per QALY be used.

Figure 39: ICERs by age and IVF cycle for unexplained cause using the IVF Predict and Hunault prediction models

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£105,571	£66,099	£56,746	£52,061	£50,046	£48,370
21	£89,012	£60,105	£52,490	£48,540	£46,851	£45,420
22	£76,592	£54,953	£48,714	£45,370	£43,954	£42,730
23	£66,902	£50,463	£45,331	£42,491	£41,304	£40,256
24	£59,097	£46,498	£42,267	£39,852	£38,860	£37,963
25	£52,637	£42,946	£39,460	£37,408	£36,581	£35,816
26	£47,159	£39,721	£36,857	£35,116	£34,434	£33,783
27	£42,407	£36,748	£34,411	£32,941	£32,383	£31,833
28	£38,195	£33,964	£32,078	£30,845	£30,397	£29,937
29	£34,383	£31,312	£29,816	£24,032	£25,602	£23,899
30	£30,862	£28,744	£24,328	£22,777	£24,296	£22,735
31	£27,712	£24,480	£22,810	£21,353	£22,772	£21,362
32	£24,312	£22,629	£21,206	£19,957	£21,357	£20,141
33	£21,656	£21,041	£19,822	£18,745	£20,123	£19,070
34	£19,530	£19,670	£18,619	£17,686	£19,041	£18,125
35	£29,449	£28,898	£27,290	£25,883	£28,463	£27,013
36	£26,428	£27,060	£25,675	£24,458	£26,947	£25,689
37	£23,968	£25,444	£24,245	£23,188	£25,592	£24,497
38	£21,292	£23,371	£22,367	£21,476	£23,720	£22,798
39	£19,604	£22,107	£21,232	£20,453	£22,625	£21,818
40	£55,452	£55,889	£55,057	£53,995	£54,558	£54,536
41	£48,091	£50,002	£49,902	£49,324	£50,042	£50,204
42	£42,094	£44,935	£45,360	£45,146	£45,971	£46,272
43	£279,319	£248,700	£228,497	£212,654	£212,172	£207,301
44	£167,711	£177,531	£174,612	£168,918	£171,423	£170,177
45	£1,159,761	£1,529,211	£1,312,835	£1,060,355	£1,044,404	£944,048

Figure 40: Cost-effectiveness of IVF by age and IVF cycles for unexplained cause using IVF Predict and Hunault prediction models



Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

Tubal cause

It was assumed that spontaneous conception leading to live birth was not possible for tubal causes of infertility and therefore no EM prediction model was required for this analysis. In this sensitivity analysis 6 cycles of IVF are cost-effective for women aged 39 years or younger at a cost-effectiveness threshold of £20,000 per QALY. For women aged 40-42 years a single cycle of IVF is cost-effective at a cost-effectiveness threshold of £30,000 per QALY. These results are depicted in tabular and graphical format in Figure 41 and Figure 42.

Figure 41: ICERs by age and IVF cycle for tubal cause using the IVF Predict model

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£6,782	£8,723	£8,723	£8,723	£9,757	£9,757
21	£6,812	£8,761	£8,761	£8,761	£9,800	£9,800
22	£6,843	£8,801	£8,801	£8,801	£9,845	£9,845
23	£6,875	£8,842	£8,842	£8,842	£9,891	£9,891
24	£6,909	£8,886	£8,886	£8,886	£9,940	£9,940
25	£6,944	£8,931	£8,931	£8,931	£9,991	£9,991
26	£6,981	£8,979	£8,979	£8,979	£10,044	£10,044
27	£7,020	£9,028	£9,028	£9,028	£10,099	£10,099
28	£7,060	£9,081	£9,081	£9,081	£10,158	£10,158
29	£7,103	£9,135	£9,135	£9,135	£10,219	£10,219
30	£7,147	£9,192	£9,192	£9,192	£10,283	£10,283
31	£7,194	£9,252	£9,252	£9,252	£10,350	£10,350
32	£7,242	£9,315	£9,315	£9,315	£10,420	£10,420
33	£7,294	£9,381	£9,381	£9,381	£10,493	£10,493
34	£7,347	£9,450	£9,450	£9,450	£10,571	£10,571
35	£10,492	£13,924	£13,924	£13,924	£15,754	£15,754
36	£10,577	£14,036	£14,036	£14,036	£15,881	£15,881
37	£10,665	£14,154	£14,154	£14,154	£16,014	£16,014
38	£10,532	£13,955	£13,955	£13,955	£15,780	£15,780
39	£10,629	£14,083	£14,083	£14,083	£15,924	£15,924
40	£22,845	£31,483	£31,483	£31,483	£32,407	£32,991
41	£23,073	£31,797	£31,797	£31,797	£32,730	£33,320
42	£23,313	£32,129	£32,129	£32,129	£33,072	£33,668
43	£59,383	£69,178	£73,239	£75,460	£78,684	£80,999
44	£60,061	£69,968	£74,076	£76,322	£79,583	£81,925
45	£109,693	£128,293	£136,017	£140,246	£146,446	£150,904

Figure 42: Cost-effectiveness of IVF by age and IVF cycles for tubal cause using the IVF Predict prediction models

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

2

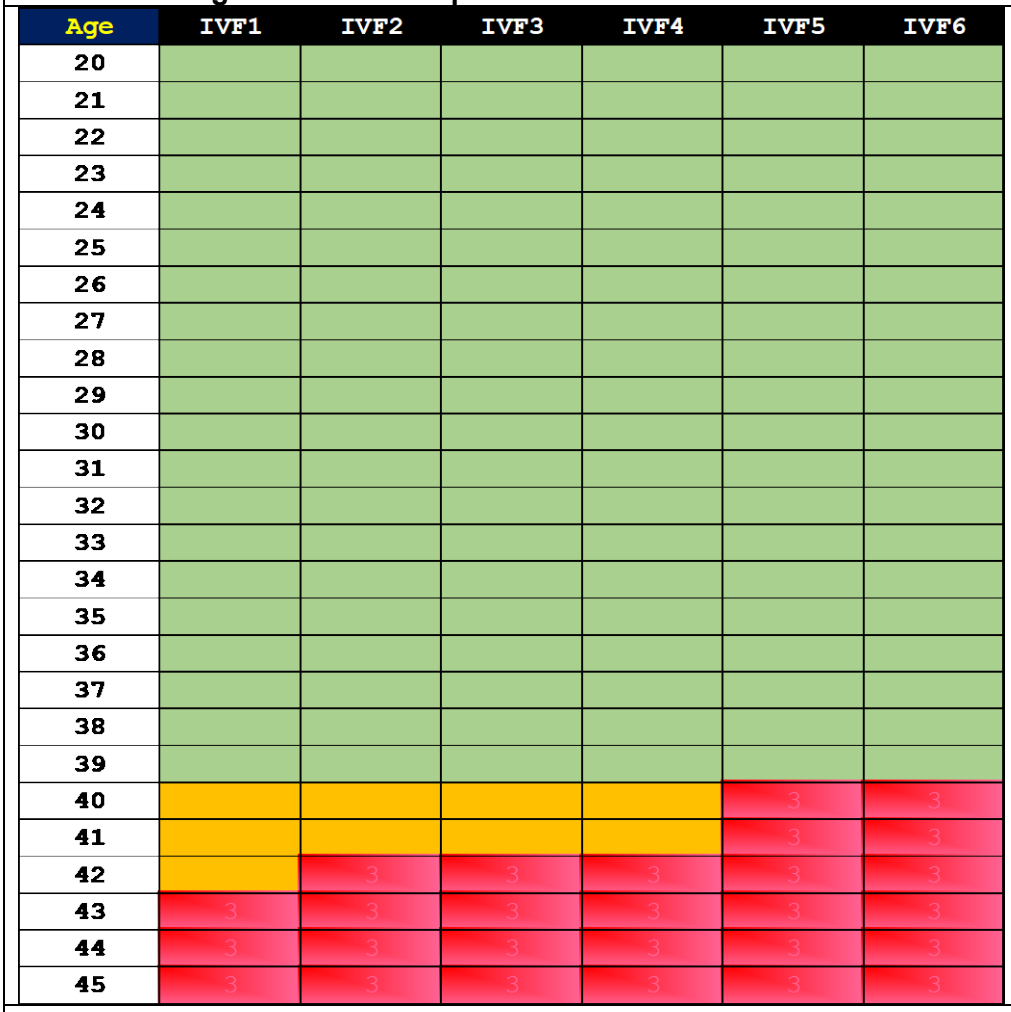
Anovulatory cause

It was assumed that spontaneous conception leading to live birth was not possible for anovulatory causes of infertility and therefore no EM prediction model was required for this analysis. As is shown in Figure 43 and , 6 cycles of IVF can be considered cost-effective for women aged 39 years and younger in this sensitivity analysis. Some cycles of IVF may also be cost-effective between the ages of 40-42 when using a cost-effectiveness threshold of £30,000 per QALY.

Figure 43: ICERs by age and IVF cycle for anovulatory cause using the IVF Predict model

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£6,477	£8,288	£8,288	£8,288	£9,253	£9,253
21	£6,505	£8,324	£8,324	£8,324	£9,294	£9,294
22	£6,535	£8,362	£8,362	£8,362	£9,336	£9,336
23	£6,566	£8,401	£8,401	£8,401	£9,380	£9,380
24	£6,598	£8,443	£8,443	£8,443	£9,426	£9,426
25	£6,632	£8,486	£8,486	£8,486	£9,474	£9,474
26	£6,667	£8,531	£8,531	£8,531	£9,525	£9,525
27	£6,704	£8,578	£8,578	£8,578	£9,578	£9,578
28	£6,742	£8,628	£8,628	£8,628	£9,633	£9,633
29	£6,783	£8,680	£8,680	£8,680	£9,691	£9,691
30	£6,825	£8,734	£8,734	£8,734	£9,751	£9,751
31	£6,870	£8,791	£8,791	£8,791	£9,815	£9,815
32	£6,916	£8,850	£8,850	£8,850	£9,881	£9,881
33	£6,965	£8,913	£8,913	£8,913	£9,951	£9,951
34	£7,017	£8,979	£8,979	£8,979	£10,025	£10,025
35	£9,953	£13,155	£13,155	£13,155	£14,863	£14,863
36	£10,033	£13,261	£13,261	£13,261	£14,982	£14,982
37	£10,117	£13,372	£13,372	£13,372	£15,107	£15,107
38	£9,994	£13,188	£13,188	£13,188	£14,891	£14,891
39	£10,086	£13,309	£13,309	£13,309	£15,027	£15,027
40	£21,486	£29,547	£29,547	£29,547	£33,844	£33,844
41	£21,701	£29,842	£29,842	£29,842	£34,182	£34,182
42	£21,927	£30,153	£30,153	£30,153	£31,026	£31,578
43	£55,587	£64,717	£68,501	£70,570	£73,569	£75,722
44	£56,222	£65,456	£69,284	£71,377	£74,410	£76,588
45	£102,537	£119,883	£127,085	£131,028	£136,804	£140,956

Figure 44: Cost-effectiveness of IVF by age and IVF cycles for anovulatory cause using the IVF Predict prediction models



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

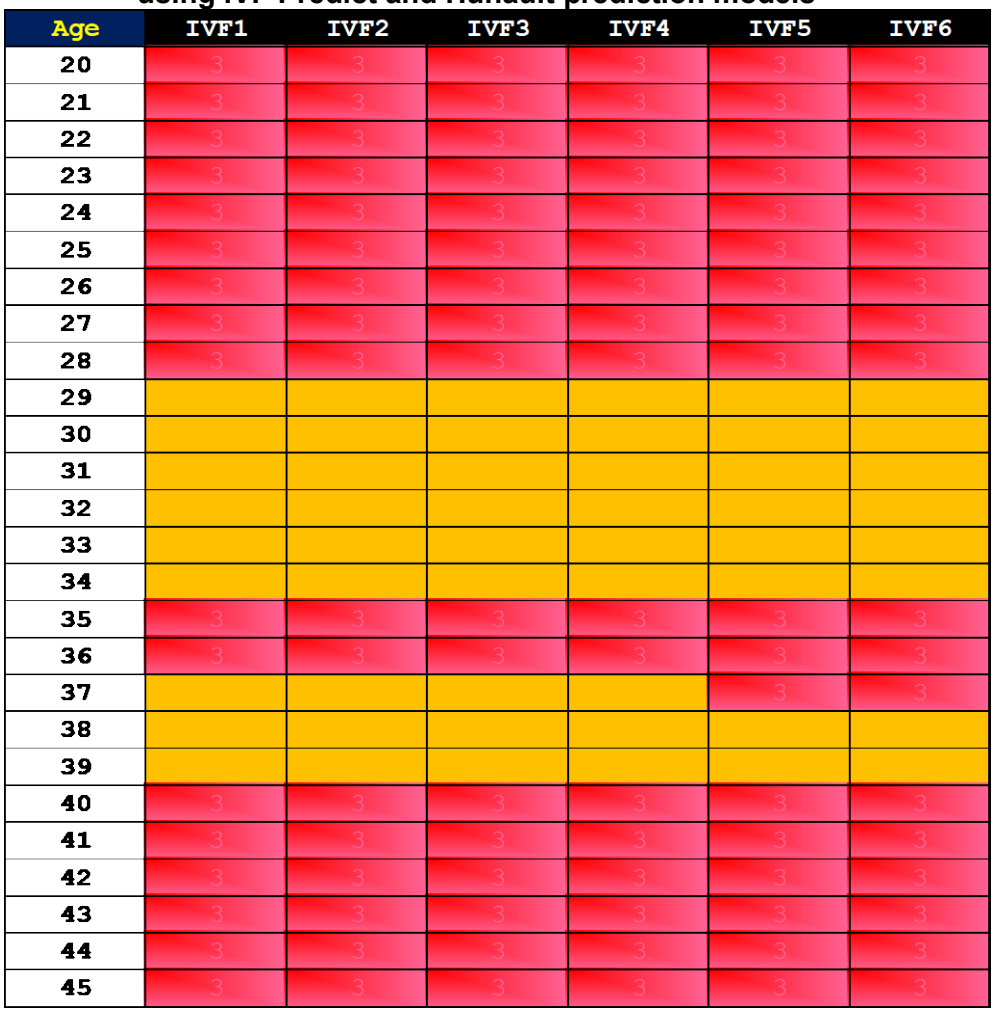
Male factor cause

In this sensitivity analysis, illustrated by Figure 45 and Figure 46, six cycles of IVF are generally cost-effective between the ages of 29-39 providing the more generous £30,000 per QALY cost-effectiveness threshold. Some seeming anomalies or exceptions in the middle of range, reflect the age banding in the IVF Predict model.

Figure 45: ICERs by age and IVF cycle for male factor cause using the IVF Predict and Hunault prediction models

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£92,780	£63,220	£55,346	£51,196	£49,582	£48,119
21	£80,517	£58,180	£51,682	£48,143	£46,793	£45,539
22	£70,897	£53,780	£48,395	£45,368	£44,242	£43,166
23	£63,124	£49,892	£45,419	£42,825	£41,890	£40,969
24	£56,683	£46,414	£42,698	£40,475	£39,703	£38,916
25	£51,225	£43,265	£40,185	£38,282	£37,651	£36,982
26	£46,504	£40,375	£37,835	£36,212	£35,704	£35,138
27	£42,337	£37,687	£35,611	£34,233	£33,833	£33,359
28	£38,587	£35,147	£33,473	£32,314	£32,010	£31,619
29	£35,146	£32,709	£31,388	£30,424	£30,205	£29,888
30	£31,928	£30,328	£29,318	£25,158	£27,349	£25,773
31	£29,056	£26,873	£25,215	£23,769	£25,809	£24,362
32	£25,819	£25,038	£23,616	£22,369	£24,350	£23,098
33	£23,260	£23,460	£22,233	£21,151	£23,079	£21,989
34	£21,191	£22,091	£21,026	£20,084	£21,962	£21,010
35	£33,685	£33,719	£33,221	£32,647	£32,825	£32,764
36	£30,430	£31,027	£30,813	£30,430	£30,681	£30,699
37	£27,773	£30,066	£29,439	£27,578	£30,797	£30,198
38	£24,770	£27,729	£26,621	£25,638	£28,639	£27,587
39	£22,951	£26,384	£25,415	£24,552	£27,455	£26,530
40	£71,546	£71,659	£70,377	£68,873	£69,686	£69,672
41	£61,777	£64,087	£63,860	£63,039	£64,063	£64,307
42	£53,973	£57,662	£58,189	£57,881	£59,053	£59,489
43	£535,468	£418,604	£365,199	£329,204	£325,397	£313,927
44	£261,778	£271,352	£262,425	£250,326	£253,261	£249,828
45	-£8,713,043	-£5,443,621	-£17,611,349	£8,758,920	£6,669,628	£3,966,485

Figure 46: Cost-effectiveness of IVF by age and IVF cycles for male factor cause using IVF Predict and Hunault prediction models



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

Mild endometriosis

In this sensitivity analysis, with the results described in Figure 47 and Figure 48, six cycles of IVF are cost-effective only between the ages of 29-39 and using a cost-effectiveness threshold of £30,000 per QALY.

Figure 47: ICERs by age and IVF cycle for mild endometriosis using the IVF Predict and Hunault prediction models

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£122,457	£71,829	£60,831	£55,450	£53,161	£51,260
21	£101,236	£65,019	£56,127	£51,609	£49,697	£48,079
22	£85,853	£59,216	£51,973	£48,163	£46,564	£45,183
23	£74,156	£54,194	£48,266	£45,042	£43,706	£42,526
24	£64,922	£49,786	£44,922	£42,189	£41,074	£40,067
25	£57,403	£45,862	£41,869	£39,552	£38,627	£37,769
26	£51,112	£42,316	£39,047	£37,088	£36,325	£35,597
27	£45,719	£39,063	£36,404	£34,753	£34,132	£33,518
28	£40,987	£36,030	£33,890	£32,510	£32,012	£31,499
29	£36,742	£33,153	£31,460	£30,319	£29,930	£25,101
30	£32,855	£30,379	£29,073	£23,868	£25,528	£23,871
31	£29,412	£25,678	£23,908	£22,367	£23,912	£22,414
32	£25,701	£23,716	£22,209	£20,889	£22,407	£21,118
33	£22,821	£22,035	£20,745	£19,608	£21,097	£19,982
34	£20,529	£20,584	£19,473	£18,490	£19,950	£18,981
35	£31,447	£31,029	£30,393	£29,771	£30,044	£29,302
36	£28,110	£28,536	£27,049	£25,745	£28,417	£27,061
37	£25,411	£26,805	£25,521	£24,389	£26,966	£25,787
38	£22,494	£24,588	£23,514	£22,563	£24,963	£23,974
39	£20,665	£23,241	£22,306	£21,475	£23,796	£22,930
40	£60,615	£60,530	£59,382	£58,082	£58,636	£58,549
41	£52,113	£53,865	£53,601	£52,874	£53,613	£53,745
42	£45,297	£48,190	£48,551	£48,252	£49,119	£49,413
43	£355,252	£298,065	£267,325	£245,100	£243,283	£236,197
44	£194,038	£202,918	£197,847	£190,056	£192,447	£190,387
45	£2,685,855	£4,341,613	£2,788,248	£1,820,885	£1,731,778	£1,453,333

Figure 48: Cost-effectiveness of IVF by age and IVF cycles for mild endometriosis using IVF Predict and Hunault prediction models

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

Combined causes

Figure 49 and Figure 50 show the results of using different model from the base case analysis for tubal cause. It generally seems to suggest that 6 cycles of IVF are cost-effective at a cost-effectiveness threshold of £30,000 QALY between the ages of 28-39.

Figure 49: ICERs by age and IVF cycle for combined causes using the IVF Predict and Hunault prediction models

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	£81,628	£58,215	£51,539	£47,939	£46,531	£45,248
21	£71,727	£53,773	£48,234	£45,153	£43,972	£42,870
22	£63,786	£49,868	£45,257	£42,613	£41,626	£40,679
23	£57,253	£46,396	£42,551	£40,280	£39,458	£38,645
24	£51,760	£43,274	£40,069	£38,117	£37,438	£36,742
25	£47,046	£40,433	£37,769	£36,094	£35,538	£34,945
26	£42,923	£37,814	£35,612	£34,179	£33,731	£33,230
27	£39,250	£35,367	£33,563	£32,345	£31,992	£31,572
28	£35,915	£33,045	£31,589	£30,562	£30,294	£29,946
29	£32,831	£30,807	£29,657	£25,091	£27,185	£25,588
30	£29,926	£26,963	£25,364	£23,942	£25,965	£24,491
31	£27,306	£25,542	£23,988	£22,629	£24,518	£23,167
32	£24,355	£23,821	£22,486	£21,313	£23,153	£21,981
33	£22,007	£22,338	£21,185	£20,167	£21,960	£20,940
34	£20,100	£21,050	£20,049	£19,161	£20,912	£20,019
35	£31,449	£31,646	£31,252	£30,760	£30,938	£30,896
36	£28,523	£30,072	£29,405	£27,424	£30,536	£29,905
37	£26,117	£28,467	£27,241	£26,158	£29,163	£28,008
38	£23,377	£26,292	£25,262	£24,347	£27,154	£26,180
39	£21,709	£25,037	£24,135	£23,331	£26,050	£25,193
40	£65,145	£65,913	£65,028	£63,824	£64,642	£64,708
41	£56,782	£59,291	£59,272	£58,638	£59,631	£59,912
42	£49,985	£53,609	£54,215	£54,015	£55,131	£55,574
43	£397,492	£338,899	£305,499	£280,836	£279,436	£271,770
44	£222,289	£233,978	£228,772	£220,136	£223,330	£221,243
45	£3,470,640	£6,110,103	£3,660,685	£2,292,567	£2,173,163	£1,802,631

Figure 50: Cost-effectiveness of IVF by age and IVF cycles for combined causes using IVF Predict and Hunault prediction models

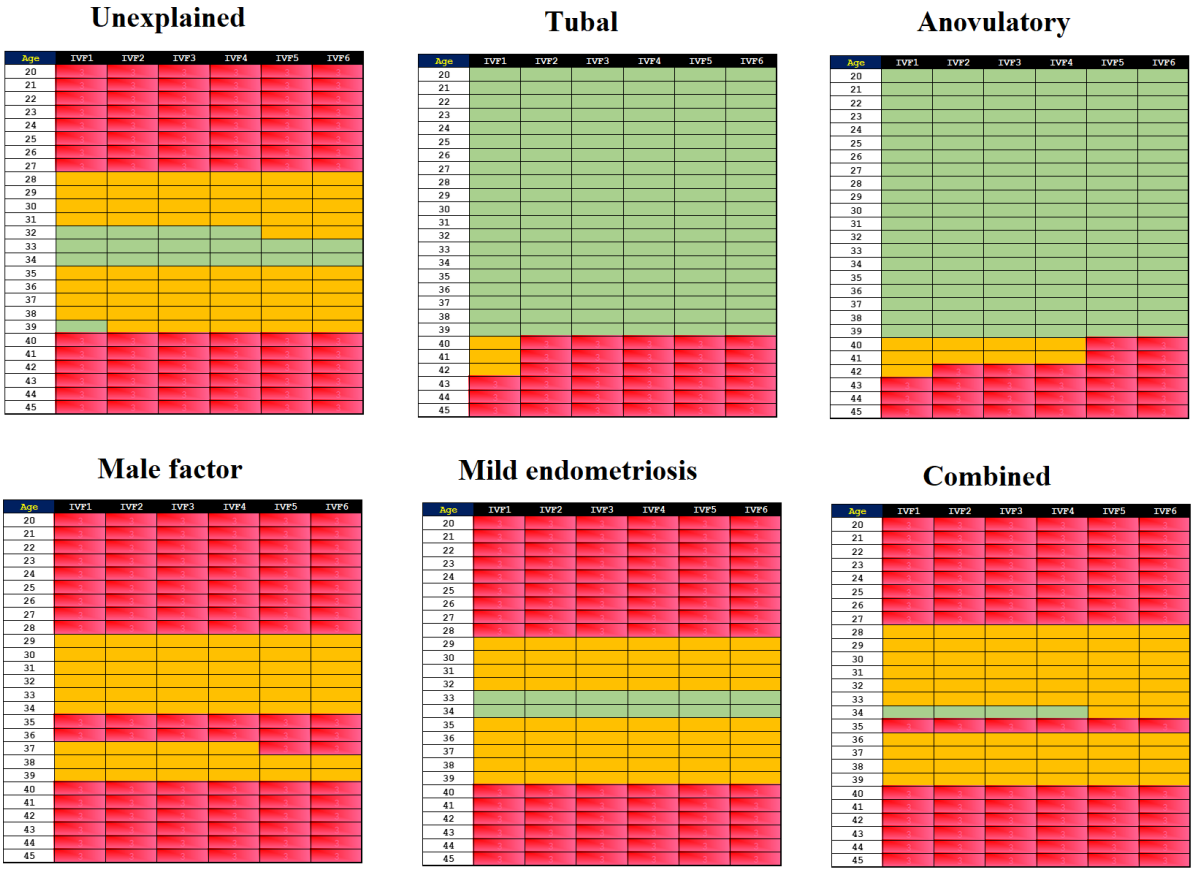
Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

Figure 52 summarises the cost-effectiveness for different causes of subfertility using the IVF Predict and Hunault prediction models.

Figure 51: Summary charts of cost-effectiveness of IVF cycles by age for different causes of subfertility using the IVF Predict and, where applicable, the Hunault prediction models



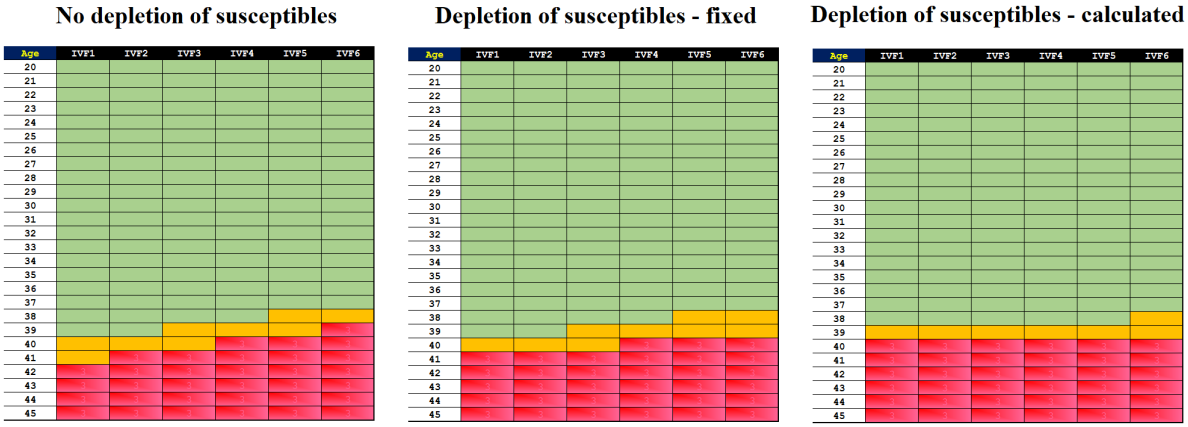
Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Depletion of susceptibles

Unexplained cause

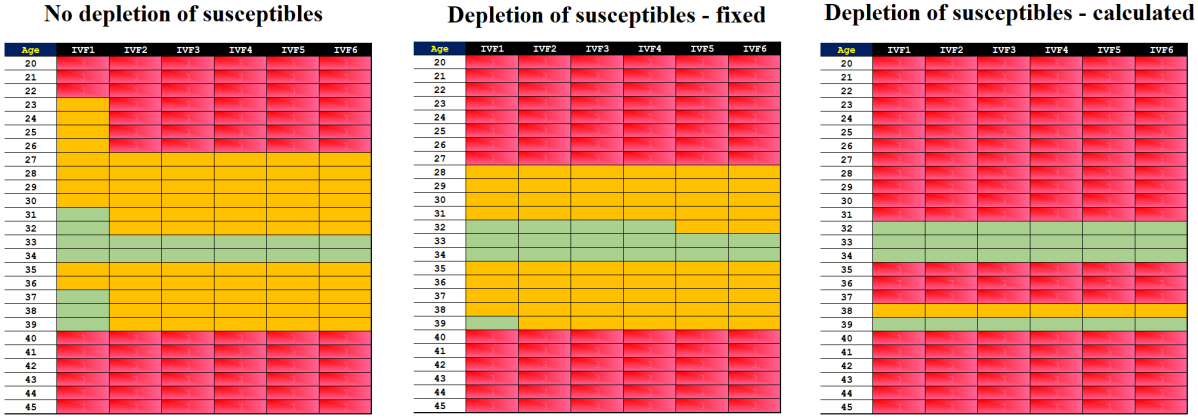
Figure 52 and Figure 53 show the impact of using differing assumptions regarding depletion of susceptibles in a population with subfertility of unknown cause, trying to conceive for 2 years prior to IVF and not using ICSI. In Figure 52 van Eekelen and OPIS pre-IVF prediction models are used and Figure 53 shows the cost-effectiveness by age and number of cycles using Hunault and IVF Predict. Whilst there is small change in age thresholds with the base case prediction models a far more dramatic impact is witnessed using Hunault and IVF Predict.

Figure 52: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for different depletion of susceptible assumptions for unexplained cause using the prediction models for the base case analysis



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Figure 53: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for different depletion of susceptible assumptions for unexplained cause using the IVF Predict and Hunault prediction models



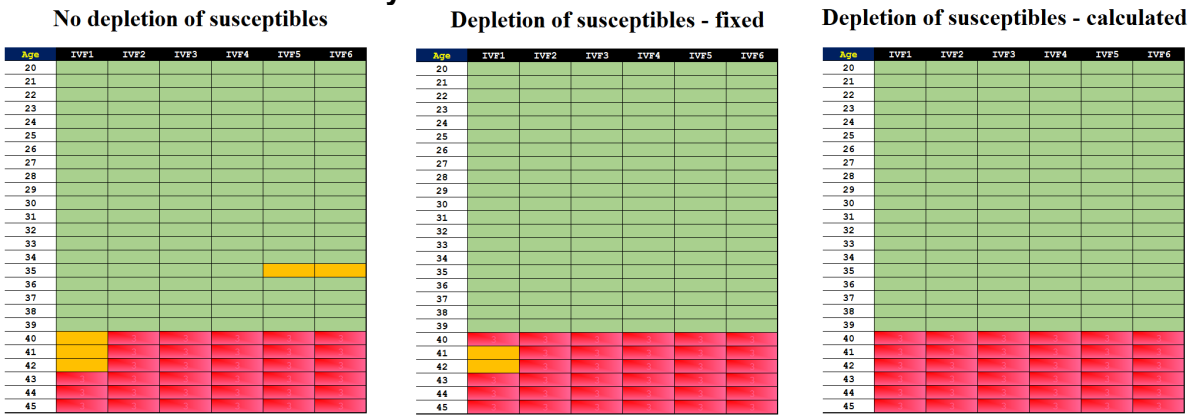
Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

Mild endometriosis cause

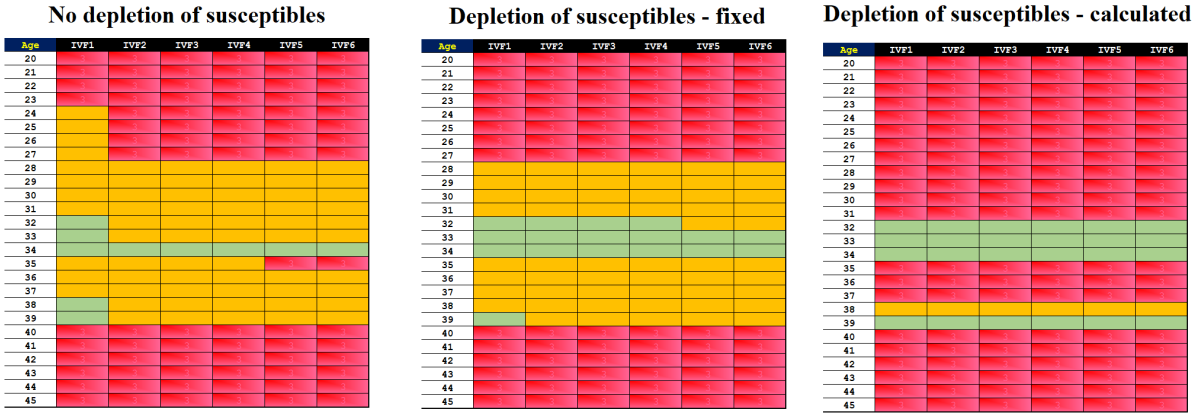
The impact of different assumptions about depletion of susceptibles for mild endometriosis cause is shown in Figure 54 and Figure 55 for different prediction models. It suggests that the base case analysis is not especially sensitive to these assumptions. However, when using the IVF Predict and Hunault prediction models, then depletion of susceptible assumptions are shown to have a dramatic impact on model conclusions.

Figure 54: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for different depletion of susceptible assumptions for mild endometriosis for the base case analysis



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Figure 55: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for different depletion of susceptible assumptions for mild endometriosis using the IVF Predict and Hunault prediction models

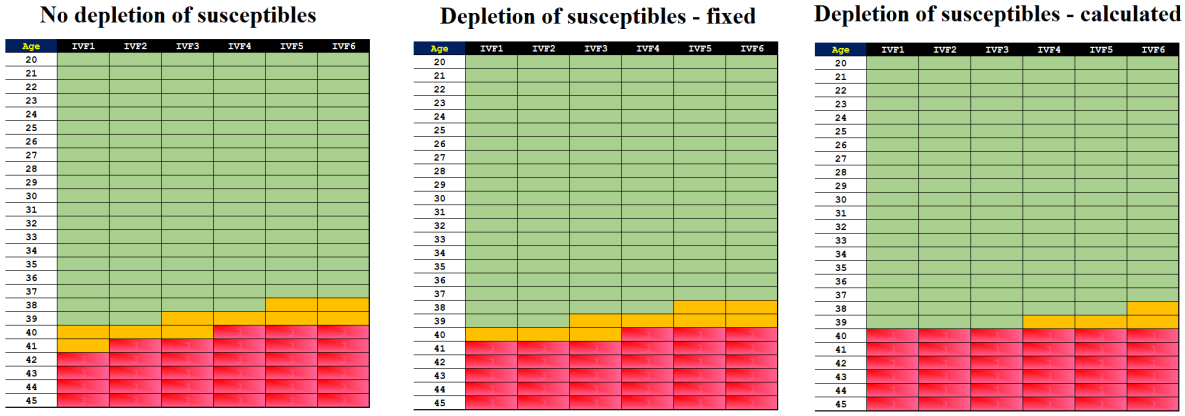


Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Male factor (no ICSI)

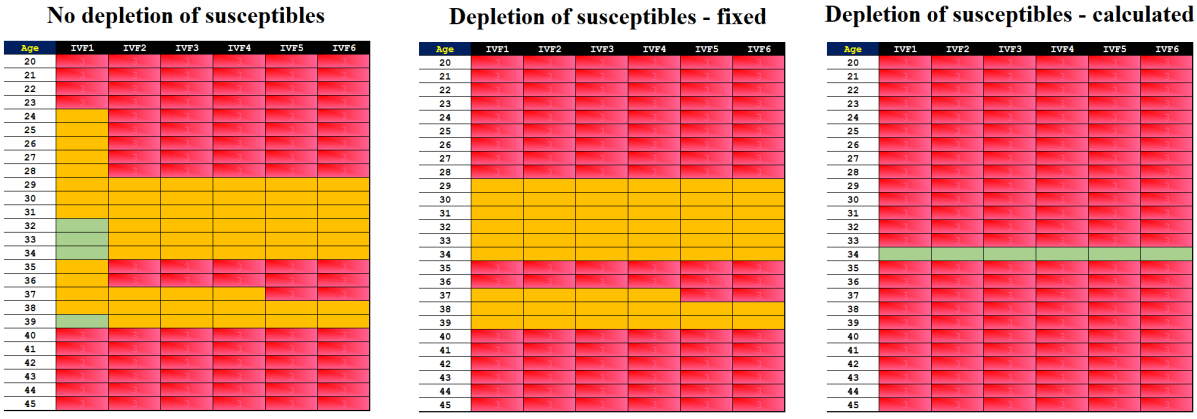
Figure 56 shows the impact of depletion of susceptible assumption on the model results for male factor cause when using the base case prediction models. A quite limited impact of these assumptions is observed in this case. However, in Figure 57 it can be seen that these assumptions fundamentally change conclusions when using Hunault and IVF Predict.

Figure 56: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for different depletion of susceptible assumptions for male factor cause for the base case analysis



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Figure 57: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for different depletion of susceptible assumptions for mild endometriosis using the IVF Predict and Hunault prediction models

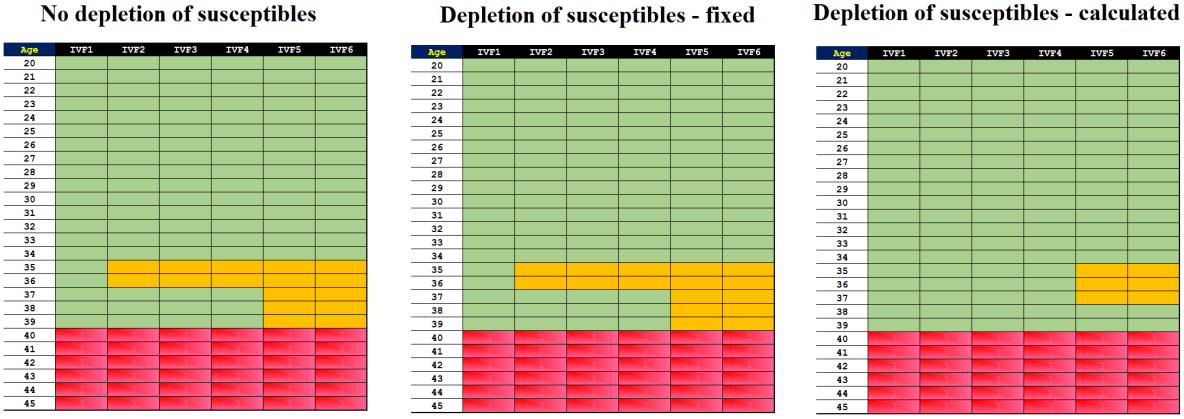


Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Combined causes

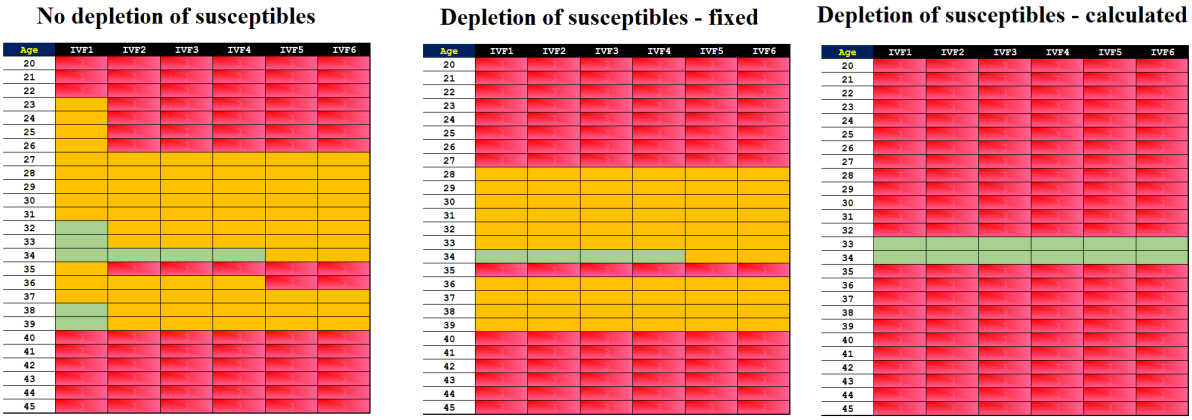
As with other causes of subfertility, Figure 58 and Figure 59 combined show that for combined causes, the importance of depletion of susceptibles assumptions depends on the choice of prediction models. If large depletion of susceptible effects are assumed, then IVF ceases to be cost-effective for analysis based on the Hunault and IVF Predict models.

Figure 58: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for different depletion of susceptible assumptions for combined causes for the base case analysis



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Figure 59: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for different depletion of susceptible assumptions for combined causes using the IVF Predict and Hunault prediction models



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Threshold analyses

Tubal

The assumption that spontaneous conception leading to live birth is not possible for tubal causes of subfertility was relaxed to ascertain the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles. This is shown for cost-effectiveness thresholds of £20,000 per QALY and £30,000 per QALY respectively in Figure 60 and Figure 61.

Figure 60: Threshold analysis indicating the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles at a cost-effectiveness threshold of £20,000 per QALY for a tubal cause of subfertility

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
21	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
22	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
23	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
24	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
25	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
26	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
27	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
28	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
29	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
30	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
31	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
32	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
33	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
34	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
35	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
36	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
37	45%	45%	45%	45%	45%	45%
38	35%	35%	35%	35%	35%	35%
39	15%	15%	15%	###	###	###
40	1%	###	###	###	###	###
41	###	###	###	###	###	###
42	###	###	###	###	###	###
43	###	###	###	###	###	###
44	###	###	###	###	###	###
45	###	###	###	###	###	###

≥50%
40%-49%
26%-39%
11%-25%
1%-10%
Never cost-effective

1

Figure 61: Threshold analysis indicating the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles at a cost-effectiveness threshold of £30,000 per QALY for a tubal cause of subfertility

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
21	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
22	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
23	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
24	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
25	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
26	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
27	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
28	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
29	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
30	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
31	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
32	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
33	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
34	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
35	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
36	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
37	48%	48%	48%	48%	48%	48%
38	36%	36%	36%	36%	36%	36%
39	22%	22%	22%	22%	22%	22%
40	8%	8%	8%	8%	###	###
41	1%	###	###	###	###	###
42	###	###	###	###	###	###
43	###	###	###	###	###	###
44	###	###	###	###	###	###
45	###	###	###	###	###	###

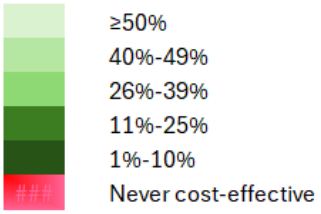
≥50%
40%-49%
26%-39%
11%-25%
1%-10%
Never cost-effective

Anovulatory

The assumption that spontaneous conception leading to live birth is not possible for tubal causes of subfertility was relaxed to ascertain the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles. This is shown for cost-effectiveness thresholds of £20,000 per QALY and £30,000 per QALY respectively in Figure 62 and Figure 63.

Figure 62: Threshold analysis indicating the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles at a cost-effectiveness threshold of £20,000 per QALY for an anovulatory cause of subfertility

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
21	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
22	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
23	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
24	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
25	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
26	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
27	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
28	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
29	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
30	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
31	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
32	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
33	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
34	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
35	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
36	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
37	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
38	44%	44%	44%	44%	44%	44%
39	29%	29%	29%	29%	29%	300%
40	7%	7%	300%	300%	300%	300%
41	300%	300%	300%	300%	300%	300%
42	300%	300%	300%	300%	300%	300%
43	300%	300%	300%	300%	300%	300%
44	300%	300%	300%	300%	300%	300%
45	300%	300%	300%	300%	300%	300%



1

Figure 63: Threshold analysis indicating the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles at a cost-effectiveness threshold of £30,000 per QALY for an anovulatory cause of subfertility

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
21	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
22	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
23	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
24	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
25	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
26	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
27	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
28	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
29	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
30	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
31	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
32	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
33	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
34	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
35	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
36	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
37	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
38	47%	47%	47%	47%	47%	47%
39	32%	32%	32%	32%	32%	32%
40	16%	16%	16%	16%	16%	16%
41	4%	4%	4%	###	###	###
42	0%	###	###	###	###	###
43	###	###	###	###	###	###
44	###	###	###	###	###	###
45	###	###	###	###	###	###

≥50%
40%-49%
26%-39%
11%-25%
1%-10%
Never cost-effective

Severe endometriosis

The assumption that spontaneous conception leading to live birth is not possible for severe endometriosis causes of subfertility was relaxed to ascertain the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles. This is shown for cost-effectiveness thresholds of £20,000 per QALY and £30,000 per QALY respectively in Figure 64 and Figure 65.

Figure 64: Threshold analysis indicating the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles at a cost-effectiveness threshold of £20,000 per QALY for severe endometriosis cause of subfertility

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
21	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
22	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
23	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
24	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
25	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
26	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
27	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
28	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
29	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
30	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
31	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
32	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
33	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
34	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
35	39%	39%	39%	39%	39%	39%
36	39%	39%	39%	39%	39%	39%
37	38%	38%	38%	38%	38%	38%
38	39%	39%	39%	39%	39%	39%
39	39%	39%	39%	39%	39%	39%
40	300%	300%	300%	300%	300%	300%
41	300%	300%	300%	300%	300%	300%
42	300%	300%	300%	300%	300%	300%
43	300%	300%	300%	300%	300%	300%
44	300%	300%	300%	300%	300%	300%
45	300%	300%	300%	300%	300%	300%

≥50%
40%-49%
26%-39%
11%-25%
1%-10%
Never cost-effective

Figure 65: Threshold analysis indicating the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles at a cost-effectiveness threshold of £30,000 per QALY for severe endometriosis cause of subfertility

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
21	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
22	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
23	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
24	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
25	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
26	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
27	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
28	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
29	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
30	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
31	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
32	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
33	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
34	≥50%	≥50%	≥50%	≥50%	≥50%	≥50%
35	42%	42%	42%	42%	42%	42%
36	42%	42%	42%	42%	42%	42%
37	41%	41%	41%	41%	41%	41%
38	42%	42%	42%	42%	42%	42%
39	42%	42%	42%	42%	42%	42%
40	3%	3%	3%	3%	###	###
41	3%	3%	3%	3%	###	###
42	3%	3%	3%	3%	###	###
43	###	###	###	###	###	###
44	###	###	###	###	###	###
45	###	###	###	###	###	###

≥50%
40%-49%
26%-39%
11%-25%
1%-10%
Never cost-effective

Cervical

The assumption that spontaneous conception leading to live birth is not possible for cervical causes of subfertility was relaxed to ascertain the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles. This is shown for cost-effectiveness thresholds of £20,000 per QALY and £30,000 per QALY respectively in Figure 66 and Figure 67.

Figure 66: Threshold analysis indicating the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles at a cost-effectiveness threshold of £20,000 per QALY for cervical cause of subfertility

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	3%	300%	300%	300%	300%	300%
21	3%	300%	300%	300%	300%	300%
22	3%	300%	300%	300%	300%	300%
23	3%	300%	300%	300%	300%	300%
24	3%	300%	300%	300%	300%	300%
25	3%	300%	300%	300%	300%	300%
26	3%	300%	300%	300%	300%	300%
27	3%	300%	300%	300%	300%	300%
28	3%	300%	300%	300%	300%	300%
29	3%	300%	300%	300%	300%	300%
30	3%	300%	300%	300%	300%	300%
31	3%	300%	300%	300%	300%	300%
32	2%	300%	300%	300%	300%	300%
33	2%	300%	300%	300%	300%	300%
34	2%	300%	300%	300%	300%	300%
35	300%	300%	300%	300%	300%	300%
36	300%	300%	300%	300%	300%	300%
37	300%	300%	300%	300%	300%	300%
38	300%	300%	300%	300%	300%	300%
39	300%	300%	300%	300%	300%	300%
40	300%	300%	300%	300%	300%	300%
41	300%	300%	300%	300%	300%	300%
42	300%	300%	300%	300%	300%	300%
43	300%	300%	300%	300%	300%	300%
44	300%	300%	300%	300%	300%	300%
45	300%	300%	300%	300%	300%	300%

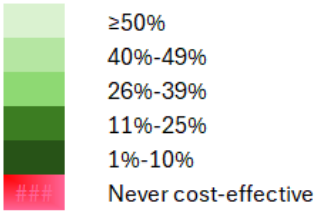


Figure 67: Threshold analysis indicating the maximum probability of spontaneous conception leading to live birth that would still be compatible with cost-effective IVF cycles at a cost-effectiveness threshold of £30,000 per QALY for cervical cause of subfertility

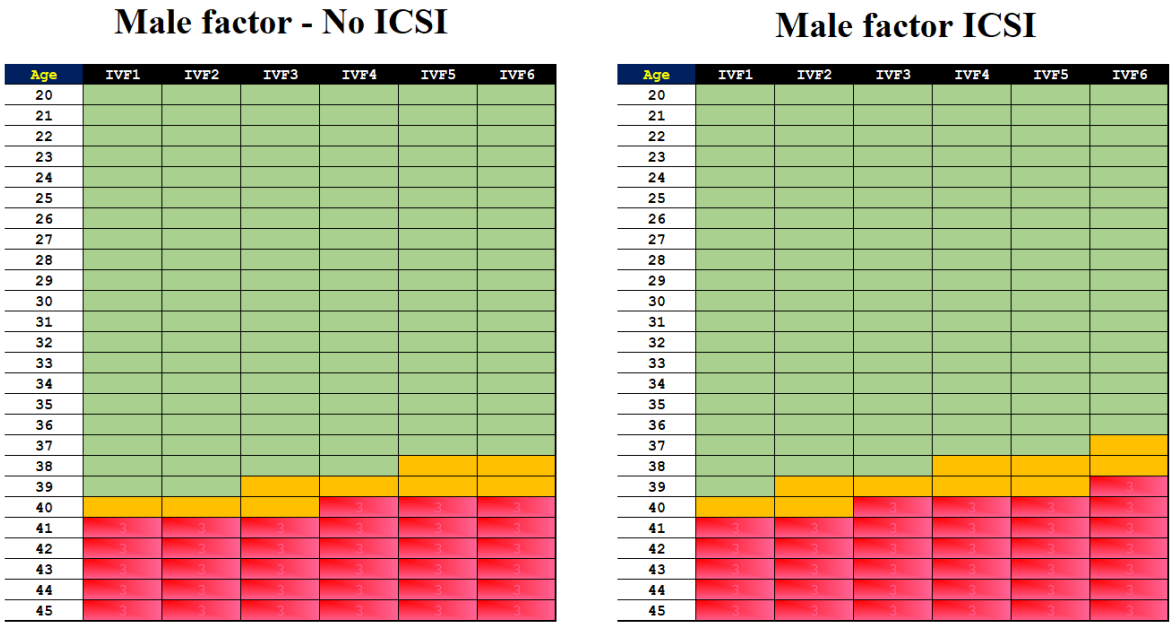
Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	18%	18%	18%	18%	18%	18%
21	18%	18%	18%	18%	18%	18%
22	18%	18%	18%	18%	18%	18%
23	17%	17%	17%	17%	17%	17%
24	17%	17%	17%	17%	17%	17%
25	17%	17%	17%	17%	17%	17%
26	17%	17%	17%	17%	17%	17%
27	16%	16%	16%	16%	16%	16%
28	16%	16%	16%	16%	16%	16%
29	16%	16%	16%	16%	16%	16%
30	16%	16%	16%	16%	16%	16%
31	15%	15%	15%	15%	15%	15%
32	15%	15%	15%	15%	15%	15%
33	15%	15%	15%	15%	15%	15%
34	14%	14%	14%	14%	14%	14%
35	1%	300%	300%	300%	300%	300%
36	1%	300%	300%	300%	300%	300%
37	1%	300%	300%	300%	300%	300%
38	1%	300%	300%	300%	300%	300%
39	1%	300%	300%	300%	300%	300%
40	300%	300%	300%	300%	300%	300%
41	300%	300%	300%	300%	300%	300%
42	300%	300%	300%	300%	300%	300%
43	300%	300%	300%	300%	300%	300%
44	300%	300%	300%	300%	300%	300%
45	300%	300%	300%	300%	300%	300%

≥50%
40%-49%
26%-39%
11%-25%
1%-10%
Never cost-effective

ICSI

Figure 64 and Figure 65 indicate the impact of using ICSI rather than IVF without ICSI. For the base case prediction models very little change in cost-effectiveness is observed. However, when using the Hunault and IVF Predict models, cost-effectiveness is found to be improved with ICSI.

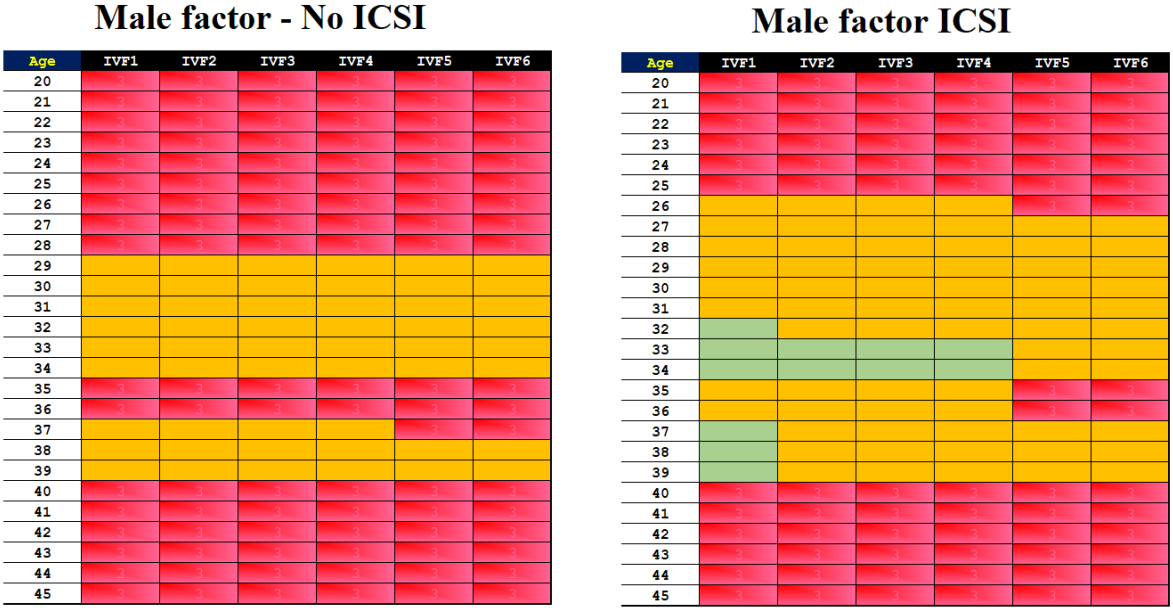
Figure 68: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for male factor cause with and without ICSI using the base case analysis prediction models



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

1

Figure 69: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for male factor cause with and without ICSI using the IVF Predict and Hunault prediction models



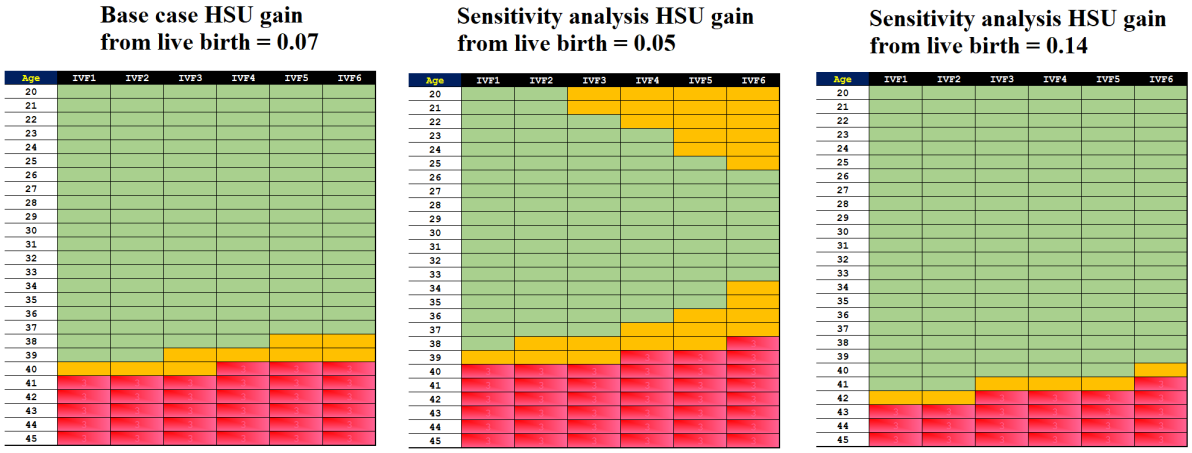
Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

2

Health state utility

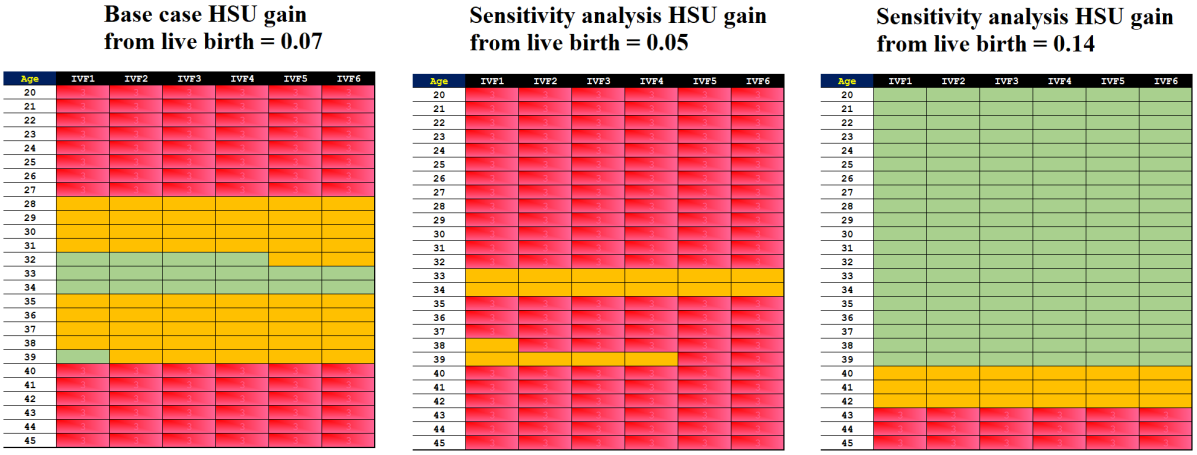
The impact of changing assumptions about health state utility gain is shown in Figure 66 and Figure 67 for the base case prediction models and for the analysis where Hunault and IVF Predict. These assumptions about health state utility have a bearing on cost-effectiveness conclusions irrespective of which prediction models are chosen but by far the greatest impact is observed with the IVF Predict and Hunault models. In this case a conservative assumption would mean that IVF would no longer be shown to be cost-effective, but conversely if greater gains are assumed then 6 cycles of IVF can be borderline cost-effective even between ages 40-42.

Figure 70: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying health state utility gain for live birth using the base case analysis prediction models



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Figure 71: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying health state utility gain using the IVF Predict and Hunault prediction models



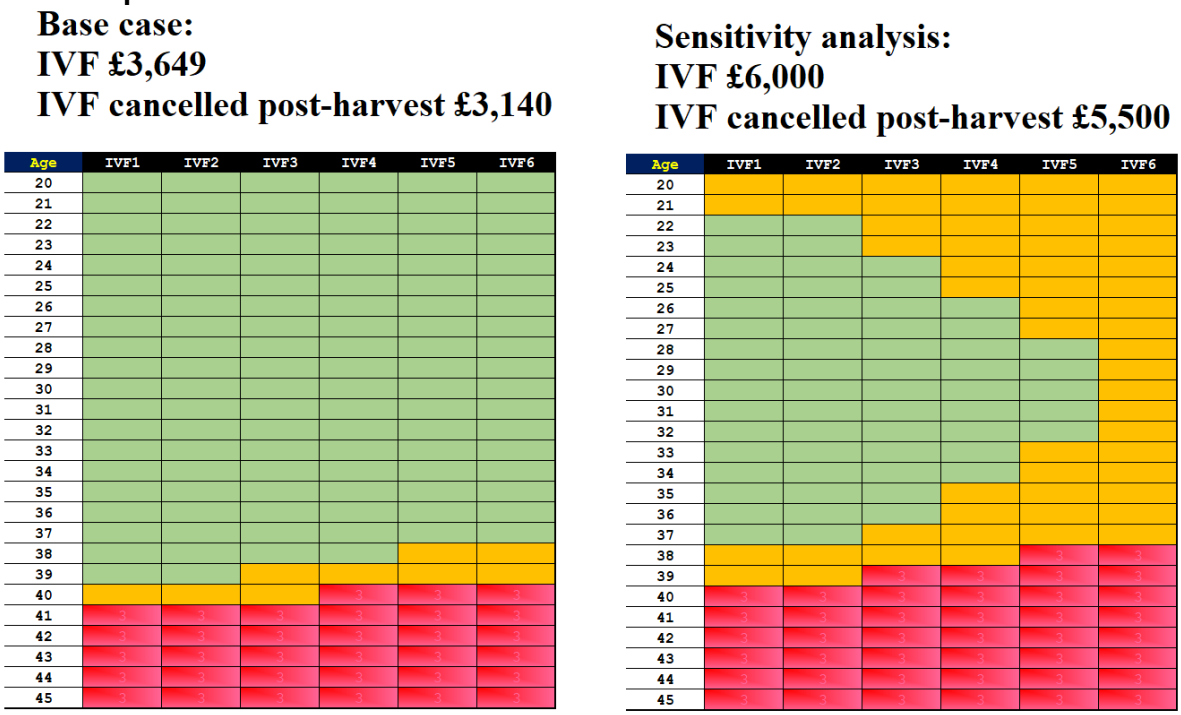
Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Cost of IVF

Whilst published costs of IVF are available, considerable uncertainty remains with respect to the actual costs to the NHS. In this sensitivity analysis, detailed in Figure 68 and Figure 69,

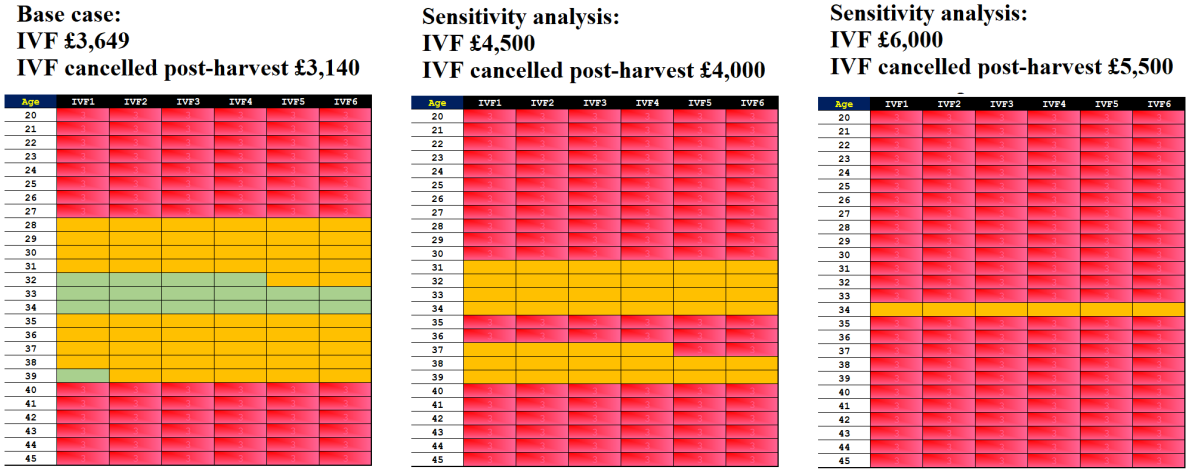
the impact of increasing the costs of IVF from their base case value is assessed. Costs of IVF are shown to be important to cost-effectiveness conclusions regardless of choice of prediction model, but results are especially sensitive to assumptions about the cost of IVF when using the Hunault and IVF Predict tools.

Figure 72: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying the cost of IVF using the base case analysis prediction models



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Figure 73: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying the cost of IVF using the IVF Predict and Hunault prediction models



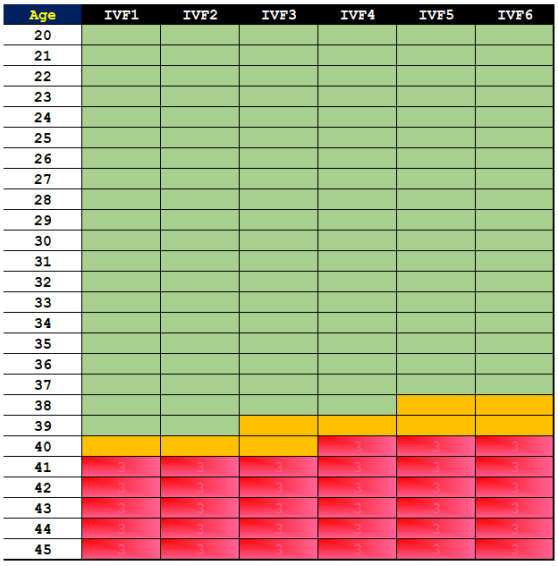
Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Spacing between IVF cycles

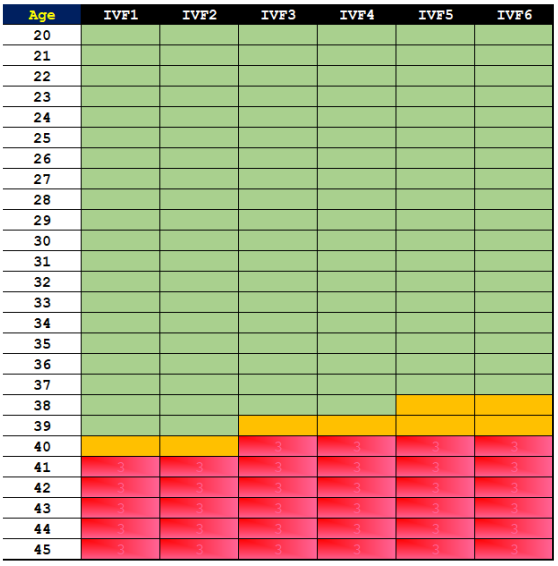
Spacing of IVF cycles is shown to have a negligible impact on cost-effectiveness conclusions when using the base case prediction models as can be seen in Figure 70. A bigger impact is observed with the Hunault and IVF Predict models (see Figure 71) although compared to other model inputs and assumptions, change to conclusions is less marked.

Figure 74: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying the spacing between IVF cycles using the base case analysis prediction models

Base case IVF spacing IVF cycles 2 months



Sensitivity analysis spacing IVF cycles 4 months



Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Figure 75: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying the spacing of IVF cycles using the IVF Predict and Hunault prediction models

Base case IVF spacing IVF cycles 2 months

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	3	3	3	3	3	3
21	3	3	3	3	3	3
22	3	3	3	3	3	3
23	3	3	3	3	3	3
24	3	3	3	3	3	3
25	3	3	3	3	3	3
26	3	3	3	3	3	3
27	3	3	3	3	3	3
28	2	2	2	2	2	2
29	2	2	2	2	2	2
30	2	2	2	2	2	2
31	2	2	2	2	2	2
32	1	1	1	1	2	2
33	1	1	1	1	1	1
34	1	1	1	1	1	1
35	2	2	2	2	2	2
36	2	2	2	2	2	2
37	2	2	2	2	2	2
38	2	2	2	2	2	2
39	1	2	2	2	2	2
40	3	3	3	3	3	3
41	3	3	3	3	3	3
42	3	3	3	3	3	3
43	3	3	3	3	3	3
44	3	3	3	3	3	3
45	3	3	3	3	3	3

Sensitivity analysis spacing IVF cycles 4 months

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	3	3	3	3	3	3
21	3	3	3	3	3	3
22	3	3	3	3	3	3
23	3	3	3	3	3	3
24	3	3	3	3	3	3
25	3	3	3	3	3	3
26	3	3	3	3	3	3
27	3	3	3	3	3	3
28	3	3	3	3	3	3
29	2	2	2	2	2	2
30	1	1	1	1	1	1
31	1	1	1	1	1	1
32	1	1	1	1	1	1
33	1	1	1	1	1	1
34	1	1	1	2	2	2
35	3	3	3	3	3	3
36	2	2	2	2	2	2
37	2	2	2	2	2	2
38	2	2	2	2	2	2
39	2	2	2	3	3	3
40	3	3	3	3	3	3
41	3	3	3	3	3	3
42	3	3	3	3	3	3
43	3	3	3	3	3	3
44	3	3	3	3	3	3
45	3	3	3	3	3	3

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Discount rate

Figure 72 shows the impact of lowering the discount rate for costs and QALYs from 3.5% to 1.5% using the base case prediction models. As would be expected this tends to improve the cost-effectiveness of IVF cycles but the overall effect is quite limited. On the other hand, as can be seen in Figure 73, the cost-effectiveness of IVF is markedly improved assuming this lower discount rate if IVF Predict and Hunault are used.

Figure 76: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying the discount rate using the base case analysis prediction models

Base case: Discount rate 3.5%

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Sensitivity analysis: Discount rate 1.5%

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Figure 77: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying the discount rate using the IVF Predict and Hunault prediction models

Base case: Discount rate 3.5%

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Sensitivity analysis: Discount rate 1.5%

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Inclusion of singleton birth costs

Figure 74 shows the impact of including singleton costs when using the base case prediction models. It can be seen that the impact on the age thresholds and number of cycles where IVF can be considered cost-effective barely changed. Figure 75 illustrates the impact when using the Hunault and IVF Predict models. As expected, inclusion of singleton costs, tends to make IVF less cost-effective, although the change to conclusions is far less than seen as a result of other changes to model assumptions.

Figure 78: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying whether singleton costs are included using the base case analysis prediction models

Base case:
Singleton costs not included

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Sensitivity analysis:
Singleton costs included

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Figure 79: Sensitivity analysis illustrating cost-effectiveness by age and IVF cycles for unexplained cause varying whether singleton costs are included using the IVF Predict and Hunault prediction models

Base case: Singleton costs not included

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	3	3	3	3	3	3
21	3	3	3	3	3	3
22	3	3	3	3	3	3
23	3	3	3	3	3	3
24	3	3	3	3	3	3
25	3	3	3	3	3	3
26	3	3	3	3	3	3
27	3	3	3	3	3	3
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40	3	3	3	3	3	3
41	3	3	3	3	3	3
42	3	3	3	3	3	3
43	3	3	3	3	3	3
44	3	3	3	3	3	3
45	3	3	3	3	3	3

Sensitivity analysis: Singleton costs included

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
20	3	3	3	3	3	3
21	3	3	3	3	3	3
22	3	3	3	3	3	3
23	3	3	3	3	3	3
24	3	3	3	3	3	3
25	3	3	3	3	3	3
26	3	3	3	3	3	3
27	3	3	3	3	3	3
28	3	3	3	3	3	3
29	3	3	3	3	3	3
30						
31						
32						
33						
34						
35	3	3	3	3	3	3
36						
37						
38						
39						
40	3	3	3	3	3	3
41	3	3	3	3	3	3
42	3	3	3	3	3	3
43	3	3	3	3	3	3
44	3	3	3	3	3	3
45	3	3	3	3	3	3

Key: green = cost-effectiveness at £20,000 per QALY cost-effectiveness threshold; amber = cost-effective at £30,000 per QALY cost-effectiveness threshold; red =not cost-effective

Discussion

The base case analyses suggested that 6 cycles of IVF were cost-effective for women aged 20-39 years for most causes of subfertility at a cost-effectiveness threshold of £30,000 per QALY, although for most causes of subfertility and at most ages 39 years and under this remained the case even using a more restrictive cost-effectiveness threshold of £20,000 per QALY. The exception was cervical cause where only 1 cycle of IVF was found to be cost-effective for women aged 35-39 using a cost-effectiveness threshold of £30,000 per QALY. Whilst 6 cycles were cost-effective for younger women, only the first cycle was cost-effective if a cost-effectiveness threshold of £20,000 per QALY was used. However, cervical causes of infertility (such as cervical tachelectomy) are extremely rare and there were less than 100 women with this diagnosis in the HFEA dataset (<0.05%) which informed IVF Predict. Therefore, the committee did not make separate recommendations for women in this subgroup as the predictions were likely to be subject to a particularly high degree of uncertainty and because the population was so small.

For women aged 40-41 years, some IVF cycles are usually cost-effective when the cause (tubal, anovulatory, severe endometriosis) is assumed to have a zero probability of spontaneous conception leading to live birth. However, the number of cost-effective cycles varies with cause, age and cost-effectiveness threshold. For unexplained and male factor cause, 3 cycles are cost-effective for women aged 40 years old providing a £30,000 per QALY threshold is used. However, for a woman aged 41, IVF was not found to be cost-effective for these groups. For mild endometriosis and combined causes, IVF was not found to be cost-effective for women aged 40-41 years although it should be noted that these analyses relied on predictions from IVF Predict as these causes were not a predictor in the OPIS pre-IVF tool.

For women aged 43 years and older, IVF was never cost-effective. This was mostly true for women aged 42 years old although the analyses suggested that 4 cycles could be cost-effective for severe endometriosis and that 1 cycle could be cost-effective for anovulatory cause providing a more permissive £30,000 per QALY cost-effectiveness threshold was used. For severe endometriosis and anovulatory causes, estimates of cost-effectiveness are improved by the simplifying assumption of a zero probability of conception with expectant management. However, the finding that 1 cycle of anovulation may be cost-effective at age 42 years is very borderline, with the threshold analysis illustrated in Figure 63 showing that this only held for a zero chance of spontaneous conception leading to live birth. For severe endometriosis, cost-effectiveness of IVF at age 42 using a £30,000 per QALY threshold depended on the probability of spontaneous conception leading to live birth being ≤3%.

Figure 31 also suggests that 1 cycle of IVF may be cost-effective for mild endometriosis at a cost-effectiveness threshold of £30,000 per QALY in women aged 42 years. However, care needs to be taken interpreting this result as it is an artefact of the age bands used in IVF Predict. Figure 31 shows that IVF is not cost-effective at ages 40-41 which are grouped in the same age 40-42 band used in the IVF Predict model. So, in IVF Predict the same prediction for live birth is made across the group when in practice, natural history dictates that effectiveness of IVF will decline across this age group. However, in the Hunault model used to estimate EM cumulative live birth rates, prediction of spontaneous conception leading to live birth does decline across this age bracket. This gives the erroneous impression that the absolute treatment effectiveness of IVF is increasing across the age band. Indeed, this is a general feature of analyses utilising IVF Predict that cost-effectiveness increases with age within any age band category as a fixed IVF birth rate is compared with an EM model where predictions of live birth fall with age.

The base case analyses utilised the van Eekelen prediction model to estimate spontaneous conception leading to live birth and where possible the OPIS pre-IVF tool for IVF leading to live birth. These 2 prediction models were preferred, as both were designed to predict cumulative live birth rates whereas IVF Predict and Hunault models had to be adapted to do this. Furthermore, they had not been developed or validated to produce estimates of cumulative live birth rates over time. However, very important limitations exist for all prediction models incorporated into this economic analysis and therefore a sensitivity analysis was undertaken where IVF Predict was used to estimate IVF live births and Hunault was used to obtain EM estimates, except where a zero probability of spontaneous conception leading to live birth was assumed.

This sensitivity analysis suggested that the cost-effectiveness of IVF was highly sensitive to the choice of prediction model with IVF generally less cost-effective than in the base case analysis. For, tubal and anovulatory causes, with zero probability of spontaneous conception leading to live birth assumed, 6 cycles of IVF continued to be cost-effective for women aged 20-39 using a £20,000 cost per QALY threshold. A fewer number of cycles could be considered cost-effective for women aged 40-42 years but only providing a £30,000 per QALY threshold was used for decision making. However, for unexplained, male factor, mild endometriosis and combined causes IVF was not cost-effective in any women aged 40 years and above. IVF was also found not to be cost-effective in women aged 20 years to 27 or 28 years in these causes. Whilst 6 cycles of IVF were generally cost-effective in women in their late twenties through to women aged 39 years for these causes, this conclusion mostly depended on a more generous £30,000 per QALY cost-effectiveness threshold being used.

Women in the datasets which informed the Hunault and van Eekelen prediction models were treatment naïve, however in our economic analysis they are also used to estimate spontaneous conception leading to live birth in women who had not had a live birth after completing their IVF cycles. These women are likely to have a worse prognosis than those in the van Eekelen and Hunault cohorts, as IVF success will to some extent reflect a better prognosis. The base case analysis did include an adaptation to account for the depletion of

susceptibles (fixed) during IVF but additionally a sensitivity analysis was undertaken which showed both the impact of relaxing the depletion of susceptibles adaptation or using a more stringent (calculated) adaptation.

Figure 52 shows a relatively minor effect of the assumptions made with respect to depletion of susceptibles for unexplained cause and using the OPIS pre-IVF and van Eekelen prediction models. Removing the adaptation for depletion of susceptibles entirely means that 1 cycle of IVF may now be considered for a woman aged 41 years using a cost-effectiveness threshold of £30,000 per QALY when compared to the base case analysis. Using a more stringent criteria to account for depletion of susceptibles results in IVF no longer being cost-effective beyond an age of 39 years. Across different causes of subfertility, it was generally the case that conclusions were not particularly sensitive to assumptions about depletion of susceptibles when using the base case prediction models.

On the other hand, Figure 53 for example, suggests that assumptions about depletion of susceptible can have a very big impact on model conclusions when using the IVF Predict and Hunault prediction models especially when a more stringent assumption is made compared to the base case analyses.

Male factor causes of subfertility can be an indication for ICSI and therefore we performed a sensitivity analysis for male factor cause both with and without ICSI. Figure 64 suggests that IVF with ICSI is similarly cost-effective to IVF without ICSI for male factor cause with the prediction models used in the base case analysis. However, it should be noted that this is not an incremental comparison between IVF with and without ICSI. Figure 65 shows the results of this sensitivity analysis using IVF Predict and Hunault prediction models. In this case ICSI improves the cost-effectiveness of IVF for male factor cause despite the increased cost.

Some commentators have questioned the usefulness and suitability of the QALY to evaluate the cost-effectiveness of fertility interventions (Keller, 2022 and Skedgel, 2023). Whilst a live birth could potentially improve health-related quality of life through an impact on mental health domains, the QALY would not capture other non-HRQoL benefits of fertility treatment such as life satisfaction and self-esteem. The extent and duration of any HRQoL improvements resulting from a live birth are highly uncertain and similar uncertainties pertain to partner utility. Therefore, sensitivity analysis was undertaken, using different values for the health state utility gain from a live birth, to address the uncertainty that exists with respect to both the method and valuation of treatment benefits. In all the analyses the assumption that the gains would be lifelong was retained, although reducing the health state utility gain can serve as a proxy for the impact of relaxing this assumption, as the base case health state utility gain experienced for a shorter period or a lower lifelong health state utility gain both result in reduced QALYs when compared to the base case analysis.

Figure 66 shows a comparison for unexplained subfertility for the base case analysis prediction models varying the health state utility gain from a live birth between 0.05 and 0.14 (a doubling from the base case which could be defended on the basis of considering partner utility). In this analysis, the conclusions are not particularly sensitive to the variations in health state utility. Using the lower bound estimate a few IVF cycles in some age groups only remain cost-effective when using a cost-effectiveness threshold of £30,000 per QALY and IVF treatment at age 40 ceases to be cost-effective. Whereas using the higher health state utility gain means that some IVF cycles could be considered at least borderline cost-effective up to age 42.

However, when a similar sensitivity analysis was undertaken for results obtained using IVF predict and Hunault for unexplained fertility, model cost-effectiveness conclusions were found to be extremely sensitive to the assumptions made with respect to health state utility gain, as illustrated in Figure 66. IVF ceases to be cost-effective in nearly all age groups when a health state utility gain of 0.05 from a live birth is used. On the other hand, 6 cycles of IVF

is cost-effective from ages 20-39 even using a cost-effectiveness threshold of £20,000 per QALY. Six cycles of IVF are also cost-effective between ages 40-42 if a higher £30,000 cost per QALY threshold is used.

Whilst the costs used for IVF were based on a published NHS source the committee thought they might underestimate the actual costs. So, a sensitivity analysis assessed the impact for subfertility of unexplained cause assuming a higher £6,000 cost for IVF and £5,500 for IVF cancelled post-harvest. As Figure 68 indicates, the impact of this substantial increase was not too marked using the base case prediction models. The age boundary for cost-effective treatment moved down to age 39 years from age 40 and fewer cycles remained cost-effective at a cost-effectiveness threshold of £20,000 per QALY. However, as can be seen in Figure 69, model conclusions were very sensitive to the IVF cost input if using the IVF Predict and Hunault prediction models. IVF no longer appeared to be a cost-effective treatment at a cost of £6,000 apart from a borderline exception for women aged 34 years. However, even using an intermediate IVF cost of £4,500, cost-effectiveness was reduced when compared to the base case result.

A default assumption in the model was to assume 2 months spacing between IVF cycles. A sensitivity analysis explored the implications of assuming a longer gap of 4 months between IVF cycles. However, as Figure 70 and Figure 71 indicate, model conclusions are not substantially changed by this increase in the spacing of IVF cycles.

In line with the NICE reference case, the model discounts costs and QALYs at an annual rate of 3.5%. However, the manual for developing NICE guidelines suggests that it can be appropriate to present the results of a sensitivity analysis where a lower discount rate of 1.5% is used. This could be reasonable in this evaluation given that the model assumes a lifelong benefit of treatment in women who are still of a fertile age. Therefore, a sensitivity analysis using the lower discount rate was carried out for unexplained subfertility. As illustrated in Figure 72, model conclusions were not found to be sensitive to the discount rate when using the base case prediction models. Where IVF Predict and Hunault prediction models were used, then the lower discount rate resulted in more cycles appearing cost-effective at a £20,000 cost per QALY cost-effectiveness threshold but there was no impact at the upper age boundary of treatment with this remaining at 39 years of age.

The final sensitivity analysis assessed the impact of including the cost of singleton births for women with unexplained fertility. Whilst these are a genuine “downstream” cost of IVF treatment, they are costs the NHS routinely covers for live births arising from spontaneous conception without recourse to any economic evaluation. Therefore, they were excluded from the base case analysis as it would disadvantage IVF access to include them when this is not a factor in determining access to NHS services for live birth arising from spontaneous conception. As Figure 74 shows, the inclusion of singleton costs has minimal impact on cost-effectiveness conclusions with the prediction models used in the base case analysis. There is a greater impact, but still relatively small, when using IVF Predict and Hunault models as can be seen in Figure 75.

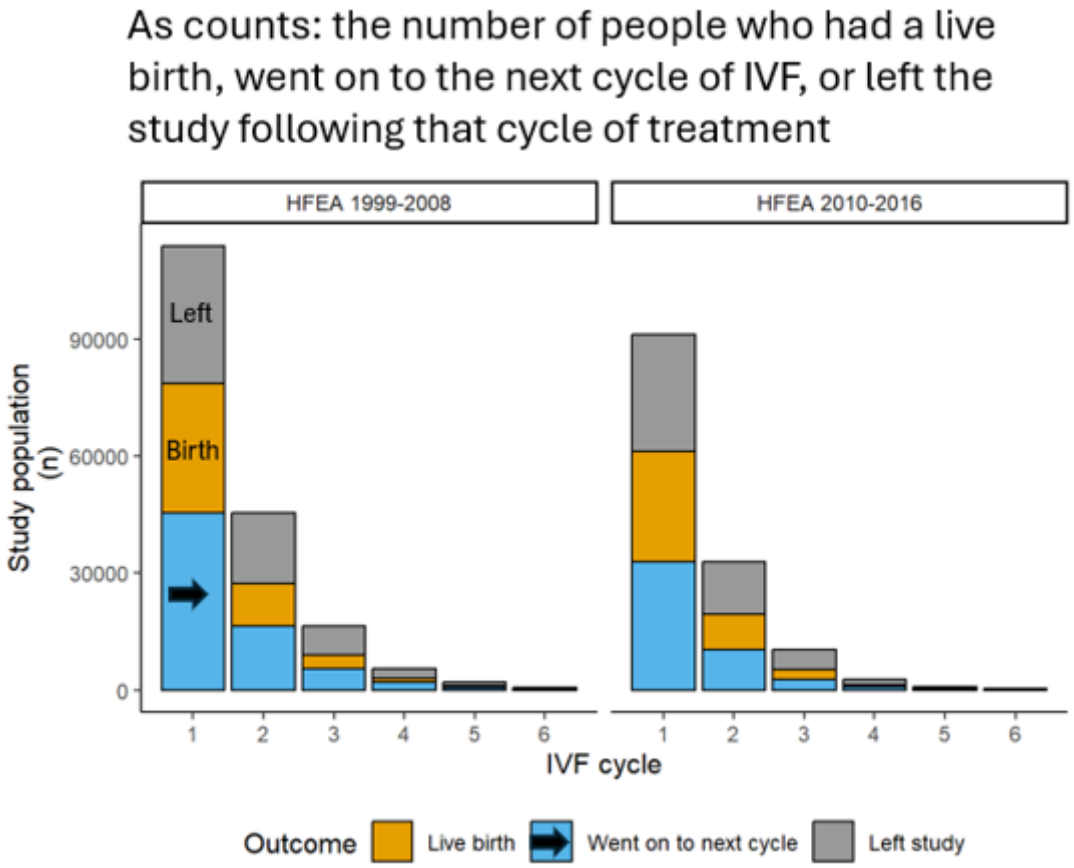
In addition to the uncertainties and modelling assumptions addressed in the sensitivity analysis, the economic evaluation has several other important limitations. In particular, any limitations in the prediction models which underpin the economic model will be propagated within this evaluation. Using the GRADE methodology the quality of the prediction model evidence was rated as very low to moderate.

Prediction models are based on observational data and as such can be prone to selection bias. This is an important limitation of this economic analysis given its reliance on these models to estimate absolute treatment effects. McLernon (2016) highlights the issue with respect to the data collection which was used to inform the development of the OPIS pre-IVF prediction tool.

> "One of the model assumptions was that those couples who discontinue treatment without a live birth still have the same chance of a live birth as those who continue, after correction for predictor effects. This assumption only applies to women with obvious barriers to continuing treatment, such as a lack of funds, divorce, or death. However, most discontinue because they have a lower chance of success. The reason for treatment withdrawal was unfortunately not available."

This discontinuation effect is illustrated in Figure 76, which shows the patients numbers in the HFEA dataset which were used to inform the OPIS pre-IVF tool predictions for each IVF cycle. Clearly people having a live birth do not progress to a subsequent IVF cycle but as shown by the grey shaded region of the bar chart, a significant proportion of those who do not have a live birth discontinue treatment. If there are systematic differences between those who continue (blue shaded region of the bar chart) and those who do not (grey) then, if these are not reflected in model predictors, observed estimates of live birth at subsequent cycles are likely to be over-estimated if continuation reflect a better chance of success.

Figure 80: HFEA cohort progression over six complete cycles

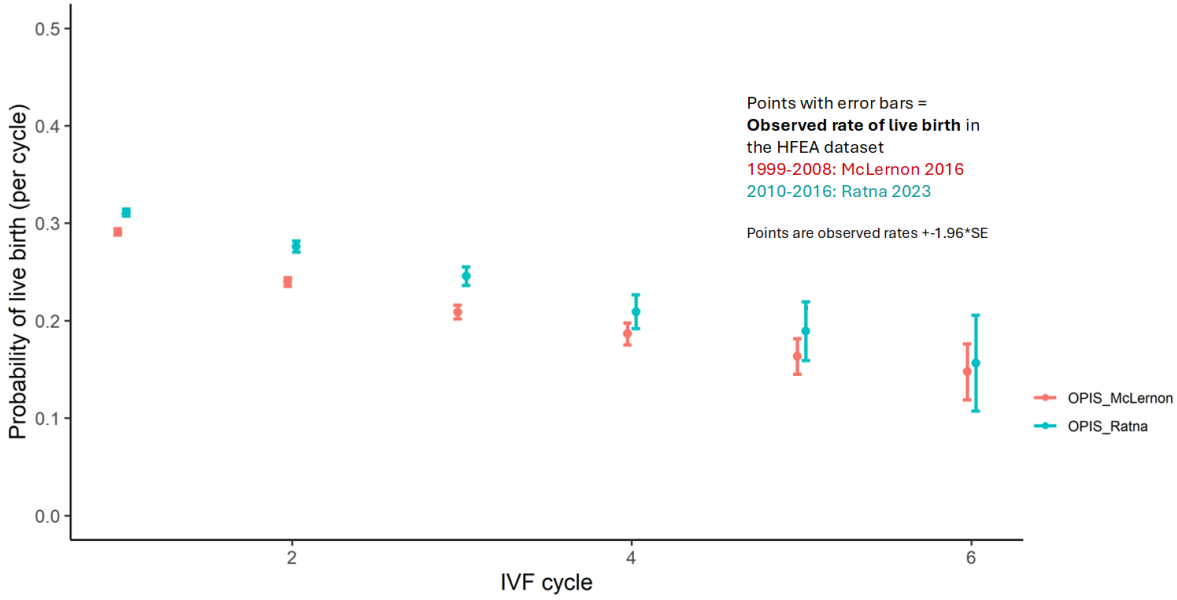


Source: NICE Technical Support Unit (TSU) using data in Ratna (2023[b])

Figure 76 also demonstrates another important limitation and that is, that predictions for higher order IVF cycles are based on observed data in a much smaller sample of patients. So, in the HFEA 2010-2016 dataset, 91,035 women had a first cycle of IVF but only 249 women in the dataset went on to have 6 cycles of IVF. The prediction models do not give confidence intervals around their estimates of the probability of live birth (which is reflected in the deterministic approach adopted in the health economic model) but small sample sizes inherently implies that there is greater uncertainty around the point estimates for predicted live birth rates for higher order IVF cycles. This is shown in Figure 77 which details

confidence intervals around the observed estimates for live birth in the HFEA 1999-2008 and 2010-2016 dataset.

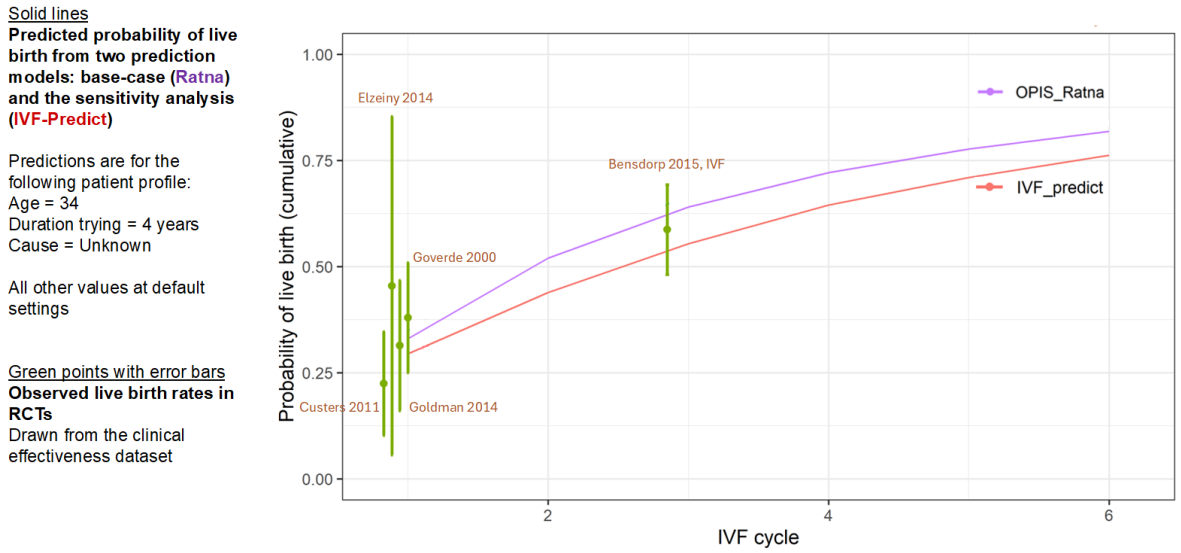
Figure 81: Error bars around observed rate of live birth in the HFEA 1999-2008 and 2010-2016 datasets



Source: NICE Technical Support Unit (TSU) using data in McLernon (2016) and Ratna (2023)

This increased uncertainty at higher order IVF cycles is of greater importance as there is a lack of randomised trials with which to compare the output of the prediction models. Figure 78 shows a comparison of the observed rates in the OPIS pre-IVF tool with RCT evidence for studies that were included for the network meta-analysis (see evidence review report K). There is an overlap between the output of the prediction models and RCTs at 1 and 3 cycles of IVF which can be interpreted as demonstrating that the prediction models are not inconsistent with RCT evidence. However, RCT data is lacking to provide similar reassurance for 4 or more IVF cycles.

Figure 82: A comparison of RCT data with the 2 IVF prediction models used in the economic analysis



Source: NICE Technical Support Unit (TSU)

There were additional limitations from using the Hunault and IVF Predict model, most importantly that neither were developed to estimate cumulative live birth rates and there has been no validation of these models as predictors of cumulative live births. The Hunault model was based on cohorts of women with a younger average age than the population covered by this economic model and the guideline committee for the previous NICE guideline (CG156) were of the view that Hunault overestimated the probability of spontaneous conception leading to live birth especially in the older age groups. Also, whilst the cohort of women included in the Hunault and van Eekelen model were attending clinics for subfertility, they may have had a better prognosis than women referred to IVF.

Conclusion

This economic analysis provides much stronger evidence for the cost-effectiveness of IVF compared with the previous NICE guideline (CG156) with lower ICERs attained in comparable analyses. Using the preferred prediction models for cumulative live birth rates from IVF and expectant management respectively, this health economic model suggested that 6 cycles of IVF was cost-effective for women aged 39 years and under, even when using the more demanding cost-effectiveness threshold of £20,000 per QALY. The only real exception was for cervical cause where only one cycle remained cost-effective for women aged 35-39 but this affects only a very small proportion of the population covered by the model and small patient numbers in the HFEA dataset make the live birth predictions more uncertain.

For women aged 40-41 years, the analysis was more equivocal, but a more limited number of IVF cycles was often cost-effective especially if inherent uncertainty in model parameters was taken into account.

None of the base case or sensitivity analyses found IVF to be cost-effective for women aged 43 and over, even when inputs were changed in a direction that favoured IVF. Many analyses also did not indicate that IVF was cost-effective for women aged 42 years even when inputs were moved in a direction favouring IVF treatment. Those few analyses which did suggest that IVF could be borderline cost-effective at age 42 years were mostly using IVF

Predict where cost-effectiveness could be an artefact of assuming the same IVF effectiveness across an age band of 40-42 years.

The model also provided some evidence that IVF with ICSI could be cost-effective for male factor cause despite the higher treatment cost.

However, model cost-effectiveness conclusions were found to be sensitive to several assumptions or inputs over which there is uncertainty. In particular, the choice of prediction model, the assumed health state utility gain from live birth, the cost of IVF and accounting for depletion of susceptibles were all found to be important cost-effective drivers in some scenarios. Furthermore, there are legitimate concerns about potential selection bias in the prediction models as well as uncertainty relating to small samples especially at 4 cycles of IVF and above.

Therefore, while the committee believed that the cost-effectiveness evidence now supported the provision of 6 cycles to women aged 39 years and below, they recognised that the evidence was not sufficiently robust to make a strong recommendation for more than 3 cycles, although they did recommend that up to 6 cycles could be considered.

The evidence for the cost-effectiveness of IVF for women aged 40-41 was less convincing but there were a number of scenarios where at least 1 cycle was cost-effective and therefore they recommended that 1 IVF cycle be offered to these women, in line with existing NICE guidance. However, the committee believed that the models no longer supported the offer of IVF to women aged 42 years.

Technical note

Most of the Results are presented as tables which detail the ICER for each successive cycle of IVF for all the ages considered in this analysis. These tabulated results are then summarised as a colour coded chart, where green shading denotes that it would be cost-effective to provide that number of IVF cycles at a more stringent cost-effectiveness threshold of £20,000 per QALY. Yellow shading denotes that it would be cost-effective to provide that number of cycles if a more permissive cost-effectiveness threshold of £30,000 per QALY is used for decision-making. Red shading mean that number of IVF cycles are not cost-effective.

Incremental analysis, in the first instance, involves assessing whether the additional benefit of providing 1 more cycle of IVF is worth the additional cost, which is assessed by calculating an ICER (as shown below) and comparing that figure with the cost-effectiveness threshold for decision making:

$$ICER_{\text{IVF cycle n}} = (Costs_{\text{IVF cycle n}} - Costs_{\text{IVF cycle n-1}}) \div (QALYs_{\text{IVF cycle n}} - QALYs_{\text{IVF cycle n-1}})$$

However, if IVF cycle $_{n-1}$ is not cost-effective, then the appropriate incremental analysis is against the next strategy which could be considered cost-effective (which could be an IVF strategy with n-2 or less cycles or expectant management, the default if no treatment is cost-effective).

Example 1

The example shown below is for women aged 40 with subfertility of unexplained cause using the base case analysis (see Figure 22, Figure 23). For convenience, the results are also repeated in Figure 79 below.

Figure 83: ICERs and cost-effectiveness of IVF cycles for a woman aged 40 years with unexplained cause of subfertility (base case analysis)

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
40	£23,052	£25,666	£29,552	£33,527	£35,680	£37,425

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
40				3	3	3

Figure 84: Cumulative live birth rate for a woman aged 40 years with unexplained cause of subfertility (base case analysis)

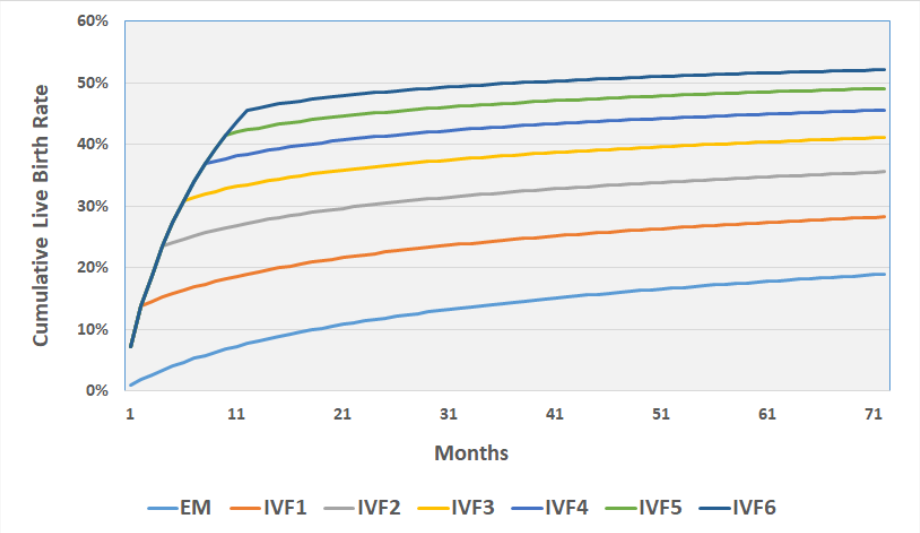


Figure 80 shows the cumulative birth rates for the different strategies that underpin these results. This shows that whilst each additional IVF cycle increases the cumulative live birth rate, it does so by a diminishing amount, and this produces ICERs which increase with each additional cycle. Table 15 shows the ICERs that would be obtained when calculating the additional costs and benefits of 1 more cycle of IVF (before ruling out for incremental comparison those strategies that are not cost-effective, see Table 16). Up to IVF3, the ICERs for all strategies fall below £30,000 per QALY and therefore can be considered cost-effective at a £30,000 per QALY cost-effectiveness threshold. However, further IVF cycles all have ICERs that are greater than £30,000 per QALY.

Table 15: ICERs calculated relative to next most effective strategy (woman aged 40 years with unexplained cause of subfertility)

Strategy	CLBR	Costs	QALYs	Inc costs	Inc QALYs	ICER
EM	0.19	£25	0.29	N/A	N/A	N/A
IVF1	0.28	£3,583	0.44	£3,559	0.15	£23,052
IVF2	0.36	£6,641	0.56	£3,058	0.12	£25,666
IVF3	0.41	£9,348	0.65	£2,706	0.09	£29,552
IVF4	0.46	£11,785	0.72	£2,437	0.07	£33,527
IVF5	0.49	£14,009	0.78	£2,224	0.06	£38,380
IVF6	0.52	£16,065	0.83	£2,055	0.05	£42,094

CLBR = Cumulative live birth rate; Inc = Incremental

However, as noted above, incremental analysis requires that more expensive strategies are compared with the next best alternative that would be considered cost-effective (if further cycles were not an option) rather than simply the next alternative, which may not be cost-effective. IVF4 is not the relevant comparator for IVF5 and therefore, the ICERs in Table 15 for IVF5 and IVF6 are different from those reported in Figure 79 and Table 16. As the ICER for IVF4 is above £30,000 per QALY (not cost-effective), IVF5 should be compared with ICER3, as shown in Table 16. This produces a lower ICER for IVF5 than reported in Table 15, but as it is still above £30,000 per QALY, the ICER for IVF6 should also be calculated relative to IVF3. The values in Table 16 correspond with those reported in Figure 79. However, as the ICERs for IVF5 and IVF6 remain above £30,000 per QALY they are still not considered cost-effective for the NHS using a £30,000 per QALY threshold.

Table 16: ICERs calculated relative to next most cost-effective strategy (woman aged 40 years with unexplained cause of subfertility)

Strategy	CLBR	Costs	QALYs	Inc costs	Inc QALYs	ICER
IVF3	0.41	£9,348	0.65	£2,706	0.09	£29,552
IVF5	0.49	£14,009	0.78	£4,662	0.13	£35,680
IVF6	0.52	£16,065	0.83	£6,717	0.18	£37,425

CLBR = Cumulative live birth rate; Inc = Incremental

Example 2

However, in some scenarios, ICERs do not increase with increase with each increasing IVF cycle. For example, for women aged 28 with subfertility of unexplained cause but using the IVF predict and Hunault prediction models (see Figure 39 and Figure 40) the ICERs actually decrease with each additional cycle, implying that each cycle is becoming relatively more cost-effective. Again, for convenience, this result is summarised in Figure 81.

Figure 85: ICERs and cost-effectiveness of IVF cycles for a woman aged 28 years with unexplained cause of subfertility (IVF Predict and Hunault models)

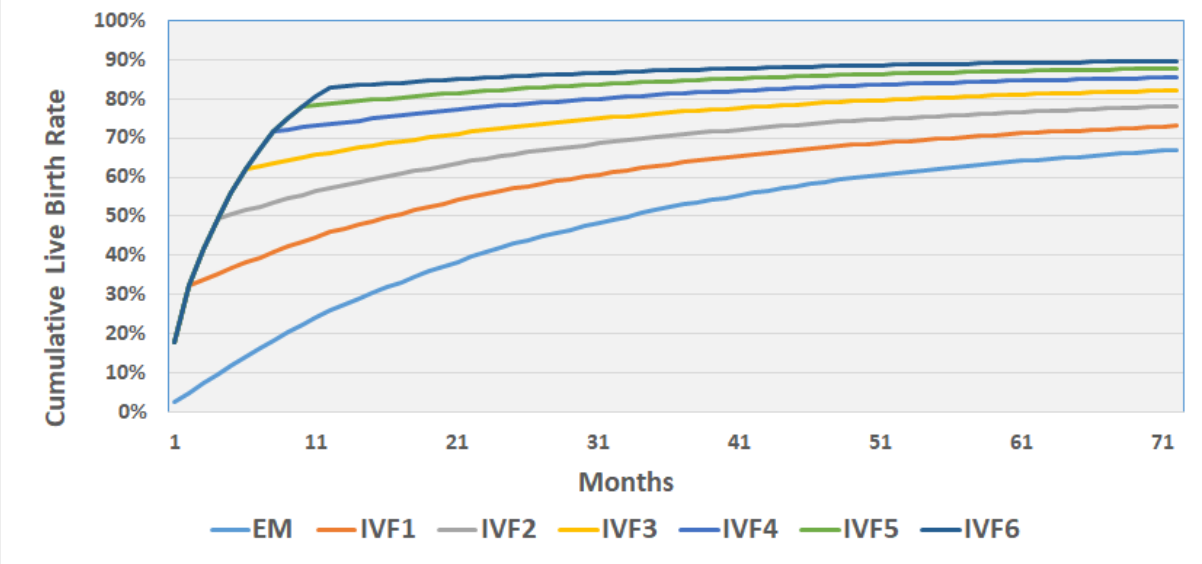
Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
28	£38,195	£33,964	£32,078	£30,845	£30,397	£29,937

Age	IVF1	IVF2	IVF3	IVF4	IVF5	IVF6
28						

Note, that in this example, all IVF cycles are colour coded as cost-effective at a cost-effectiveness threshold of £30,000 per QALY even though only IVF6 has an ICER which falls below that level. This is a convention that has been followed throughout where if IVF_n is cost-effective at a given threshold, then IVF_1 to IVF_{n-1} are also shaded as cost-effective at that threshold. IVF_1 to IVF_{n-1} are not mutually exclusive interventions to the IVF_n strategy, indeed they are subsets of it. Many women will have a live birth without reaching the maximum number of cycles and therefore we thought it more intuitive to present all cycles that are cost-effective as part of cost-effective strategy to be shown as cost-effective. However, this means that it should not be assumed that if IVF_n is cost-effective that a strategy offering less IVF cycles would be.

The cumulative live birth rates are for this scenario are shown in Figure 82.

Figure 86: Cumulative live birth rate for a woman aged 40 years with unexplained cause of subfertility (IVF Predict and Hunault models)



The first thing to note is that the gains of the first IVF cycle relative to EM are much smaller. Intuitively, therefore it might be anticipated that, given the costs of that first cycle of IVF, IVF1 might not be cost-effective. This is borne out by the ICER for IVF1 in Figure 81. However, further IVF cycles continue to increase the cumulative live birth rate, and it is possible when ICERs are decreasing with higher cycles that a point is reached where IVF is cost-effective at higher order cycles when offering fewer cycles would not be.

Table 17 shows the ICERs if calculated relative to the next IVF strategy.

Table 17: ICERs calculated relative to next most effective strategy (woman aged 40 years with unexplained cause of subfertility- IVF Predict and Hunault models)

Strategy	CLBR	Costs	QALYs	Inc costs	Inc QALYs	ICER
EM	0.75	£96	1.22	N/A	N/A	N/A
IVF1	0.78	£3,735	1.32	£3,639	0.10	£38,195
IVF2	0.82	£6,174	1.40	£2,439	0.08	£29,146
IVF3	0.85	£7,999	1.47	£1,825	0.07	£27,072
IVF4	0.88	£9,365	1.52	£1,366	0.05	£25,237
IVF5	0.90	£10,384	1.56	£1,019	0.04	£26,848
IVF6	0.91	£11,174	1.59	£790	0.03	£25,002

CLBR = Cumulative live birth rate; Inc = Incremental

These ICERs are plotted on a cost-effectiveness plane (see Figure 83). By comparing the gradient between EM and IVF1 we can see that the ICER for the first cycle of IVF relative to EM that the ICER is £38,195 (see also Table 17) which is not cost-effective at a cost-effectiveness threshold of £30,000 per QALY. Whilst IVF2 looks cost-effective relative to IVF1 with an ICER of £29,146 (see also Table 17) this is based on a comparison against an alternative that is not cost-effective. IVF2 should rather be compared against the next best alternative that would be cost-effective (EM). Recalculating the ICER for IVF2 relative to EM, we get the gradient between IVF1 and IVF2 on Figure 83, showing an ICER of £33,964 (see Figure 81) which would not be cost-effective at £30,000 per QALY. This elimination of non-

cost-effective options continues until IVF6 with the recalculated ICERs provided in Table 18 (which correspond with those in Figure 81). However, the graph also makes it clear that only IVF6 falls below the threshold of £30,000 per QALY

Figure 87: Cost-effectiveness plane showing IVF strategies for a woman aged 40 years with unexplained cause of subfertility (IVF Predict and Hunault models)

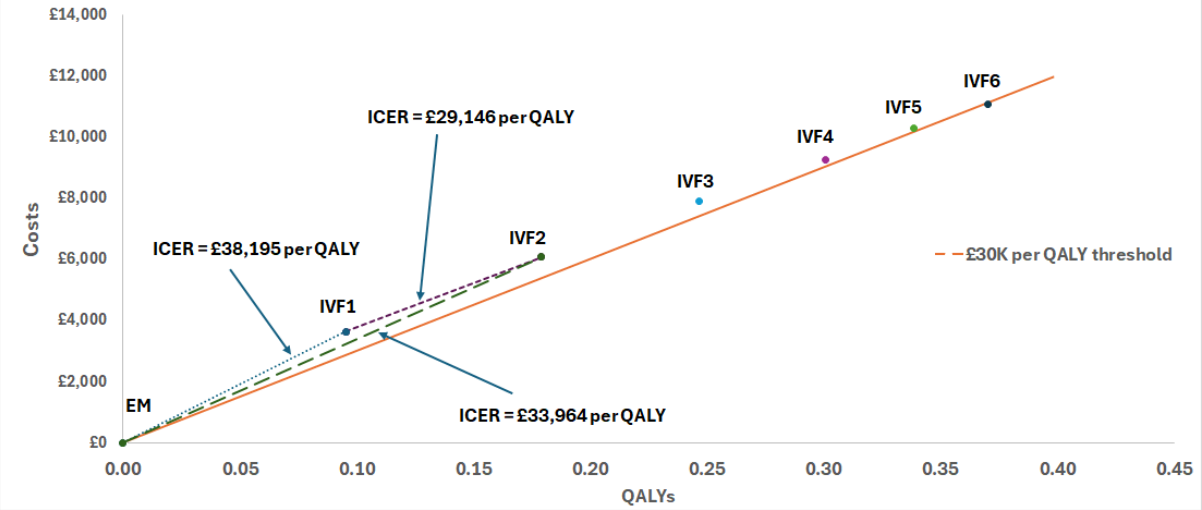


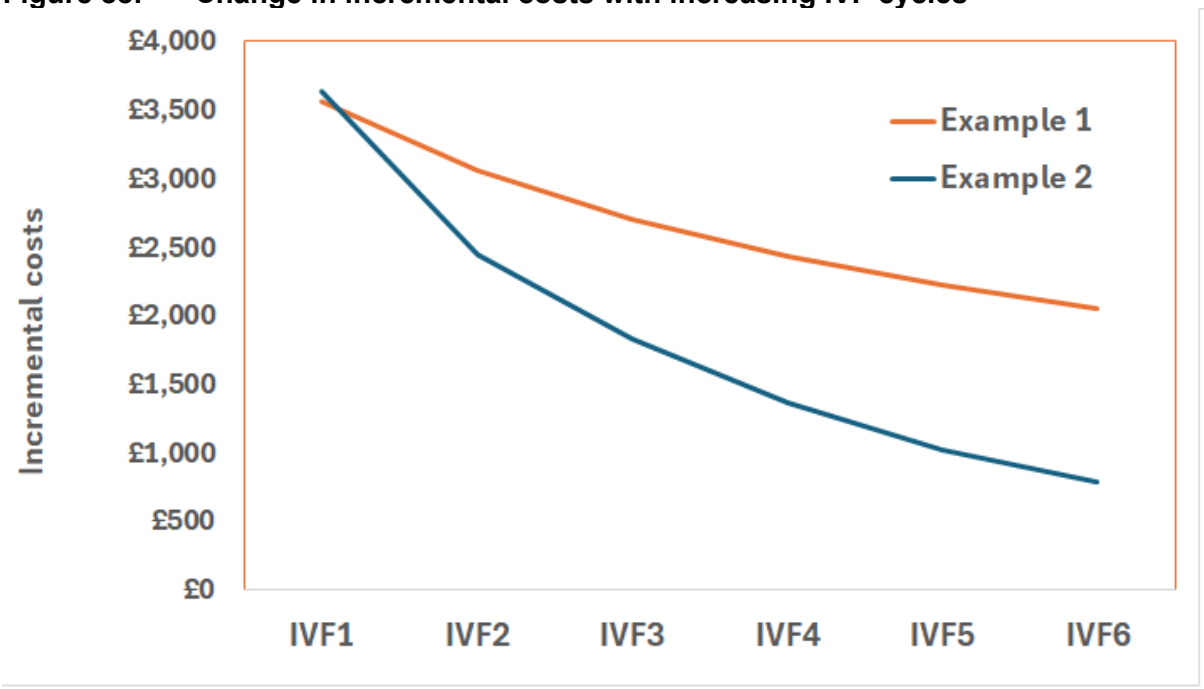
Table 18: ICERs calculated relative to next most cost-effective strategy (woman aged 40 years with unexplained cause of subfertility- IVF Predict and Hunault models)

Strategy	CLBR	Costs	QALYs	Inc costs	Inc QALYs	ICER
EM	0.75	£96	1.22	N/A	N/A	N/A
IVF1	0.78	£3,735	1.32	£3,639	0.10	£38,195
IVF2	0.82	£6,174	1.40	£6,078	0.25	£33,964
IVF3	0.85	£7,999	1.47	£7,903	0.30	£32,078
IVF4	0.88	£9,365	1.52	£9,269	0.34	£30,845
IVF5	0.90	£10,384	1.56	£10,288	0.37	£30,397
IVF6	0.91	£11,174	1.59	£11,078	0.25	£29,937

CLBR = Cumulative live birth rate; Inc = Incremental

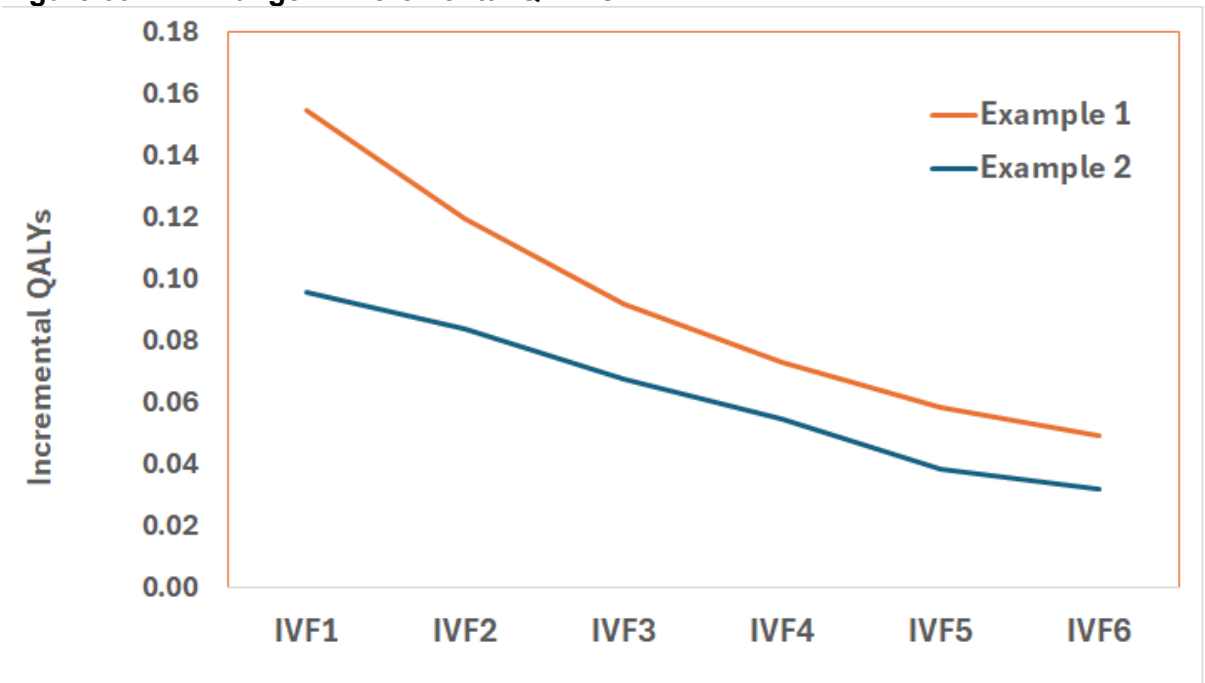
The explanation as to why the ICERs may either rise or fall with an increase in IVF cycles is the result of the interaction of the changes in costs and births that arise from increasing the number of IVF cycles. Figure 84 and Figure 85 show the change in incremental costs and incremental QALYs from example 1 (increasing ICERs) and example 2 (decreasing ICERs).

Figure 88: Change in incremental costs with increasing IVF cycles



1

Figure 89: Change in incremental QALYs



2 It is the fact that incremental costs are falling much more rapidly in example 2 (see Figure
3 84) which is driving the different patterns in ICERs with increasing numbers of IVF cycles.
4 This is because, despite the higher additional benefits of additional cycles in example 1 (see
5 Figure 85) the absolute numbers who have unsuccessful IVF (and go onto further cycles) is
6 higher because of a much lower cumulative live birth rate overall (see Figure 80 and Figure

82). In example 1 these relatively high incremental costs mean that, combined with the diminishing returns of additional cycles, that the relative cost-effectiveness decreases (increasing ICERs) with increasing IVF cycles. Whereas, in example 2, rapidly falling incremental costs mean that the relative cost-effectiveness increases (decreasing ICERs) with increasing IVF cycles, as that effect more than offsets the diminishing return of additional IVF cycles.

It should be noted that increasing cost-effectiveness with the addition of more IVF cycles is usually only consistently found when using the Hunault model to predict cumulative live birth rates from spontaneous conception. This model predicts higher cumulative live birth rates than the van Eekelen model.

References

Braakhekke 2014

Braakhekke M, Kamphuis EI, Dancet, Mol F, van der Veen F, Mol BW. Ongoing pregnancy qualifies best as the primary outcome measure of choice in trials in reproductive medicine: an opinion paper. Fertility and Sterility. 101(5):1203 – 1204, 2014

CDC 2021

Centers for Disease Control and Prevention. 2018 Assisted Reproductive Technology National Summary Report. US Dept of Health and Human Services; 2021.

Collins 1995

Collins JA, Burrows EA, Wilan AR. The prognosis for live birth among untreated infertile couples. Fertility and Sterility. 64:22-28, 1995

Cooper 2009

Cooper, Noonan, Eckardstein, Auger, Baker, Behre, Haugen, Kruger, Wang, Mbizvo, Vogelsong. World Health Organization reference values for human semen characteristics Human Reproduction. Update 16(3): 231-245, 2010

Devlin 2018

Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Economics. 27(1):7-22, 2018

Eimers 1994

Eimers JM, Velde ER, Gerritse R, Vogelzang ET, Looman CW, Habbema JD. The prediction of the chance to conceive in subfertile couples. Fertility and Sterility; 61:44-52, 1994

Gleicher 2009

Gleicher N, Barad D. Twin pregnancy, contrary to consensus, is a desirable outcome in infertility. Fertility and Sterility. 91(6):2426-31, 2009

Hunault 2004

Hunault CC, Habbema JD, Eijkemans MJ, Collins JA, Evers JL, te Velde ER. Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. Human Reproduction. 19(9):2019-26, 2004

Keller 2022

Keller E, Chambers GM. Valuing infertility treatment: Why QALYs are inadequate, and an alternative approach to cost-effectiveness thresholds. Frontiers in Medical Technology. 4:1053719, 2022

McLernon 2016

McLernon DJ, Steyerberg EW, Te Velde ER, Lee AJ, Bhattacharya S. 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.

Nelson 2011

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 Medicine. 2011 8(1): e1000386.

Ratna 2023

Ratna MB, Bhattacharya S, McLernon DJ. External validation of models for predicting cumulative live birth over multiple complete cycles of IVF treatment. Human Reproduction. 3;38(10):1998-2010, 2023 Oct. doi: 10.1093/humrep/dead165. Erratum in: Hum Reprod. 2024 May 10:deae099. doi: 10.1093/humrep/deae099. PMID: 37632223; PMCID: PMC10546080.

Ratna 2023[b]

Ratna MB, Bhattacharya S, McLernon DJ. External validation of models for predicting cumulative live birth over multiple complete cycles of IVF treatment. Human Reproduction. 38(10): 1998-2010, 2023

Scotland 2011

Scotland GS, McLernon D, Kurinczuk JJ, McNamee P, Harrild K, Lyall H, Rajkhowa M, Hamilton M, Bhattacharya S. Minimising twins in in vitro fertilisation: a modelling study assessing the costs, consequences and cost-utility of elective single versus double embryo transfer over a 20-year time horizon. British Journal of Obstetrics & Gynaecology 118(9):1073-83, 2011

Skedgel 2023

Skedgel C, Cubi-Molla P, Mott D, Gameiro S, Boivin J, Al-Janabi H, Brazier J, Markert M, Andersson FL, Jofre-Bonet M. Unmet Parenthood Goals, Health-Related Quality of Life and Apparent Irrationality: Understanding the Value of Treatments for Infertility. Pharmacoeconomics Open. 7(3):337-344, 2023

Snick 1997

Snick HK, Snick TS, Evers JL, Collins JA. The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary care study. Human Reproduction. 12:1582-1588, 1997

Sood 2020

Sood, A., Goel, A., Boda, S., & Mathur, R. Prediction of significant OHSS by ovarian reserve and ovarian response – implications for elective freeze-all strategy. Human Fertility, 25(2), 390–396, 2029

van Eekelen 2017

van Eekelen R, Scholten I, Tjon-Kon-Fat RI, van der Steeg JW, Steures P, Hompes P, van Wely M, van der Veen F, Mol BW, Eijkemans MJ, Te Velde ER, van Geloven N. Natural conception: repeated predictions over time. Human Reproduction. 32(2):346-353, 2017

Acknowledgements

We would like to acknowledge the following:

Guidelines Technical Support Unit, at University of Bristol, particularly Nicky Welton and Beatrice Downing, for providing advice and feedback.

David McLernon, Medical Statistics Team, Institute of Applied Health Sciences, for providing clarification on the OPIS pre-IVF Tool equations.

Rik van Eekelen, Amsterdam University Medical Center, for providing clarification on the van Eekelen prediction model equations.

Appendix J Excluded studies

Excluded clinical prediction model studies

Table 19: Excluded studies and reasons for their exclusion

Study	Reason
Abdelghani, A.M.A.; Hashem, M.F.A.; Aemmaar, W.A.E. (2022) Predictors of ovarian response: progress towards individualized treatment in ovulation induction and ovarian stimulation. NeuroQuantology 20(19): 3618-3627	- Outcome is not relevant
Ainsworth, Alessandra J, Barnard, Emily P, Baumgarten, Sarah C et al. (2020) Intrauterine insemination cycles: prediction of success and thresholds for poor prognosis and futile care. Journal of assisted reproduction and genetics 37(10): 2435-2442	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Al-Sabbah, S.A.S.; Abood, A.H.; Mohammed, E.A.A. (2020) The most influential factors for a successful pregnancy after in vitro fertilization by model of probit regression. Annals of Tropical Medicine and Public Health 23(14): p23142	- No calibration or discrimination data can be extracted
Alebic, Miro Simun; Stojanovic, Natasa; Zuvic-Butorac, Marta (2013) The IVF Outcome Counseling Based on the Model Combining DHEAS and Age in Patients with Low AMH Prior to the First Cycle of GnRH Antagonist Protocol of Ovarian Stimulation. International journal of endocrinology 2013: 637919	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Amini, P., Ramezanali, F., Parchehbaf-Kashani, M. et al. (2021) Factors associated with in vitro fertilization live birth outcome: A comparison of different classification methods. International Journal of Fertility and Sterility 15(2): 128-134	- Not a high-income OECD country
Arny, M and Quagliarello, J (1987) Semen quality before and after processing by a swim-up method: relationship to outcome of intrauterine insemination. Fertility and sterility 48(4): 643-8	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Arvis, P.; Lehert, P.; Guivarc'h-Leveque, A. (2019) Both high and low HCG day progesterone concentrations negatively affect live birth rates in IVF/ICSI cycles. Reproductive BioMedicine Online 39(5): 852-859	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction

Study	Reason
Arvis, P, Lesourd, F, Parneix, I et al. (2021) Long-term outcome of patients undergoing in-vitro fertilisation in France: The outcome study. Journal of gynecology obstetrics and human reproduction 50(3): 101968	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Ashrafi, M., Sadatmahalleh, S.J., Akhoond, M.R. et al. (2013) ICSI outcome in infertile couples with different causes of infertility: A cross-sectional study. International Journal of Fertility and Sterility 7(2): 88-95	- Clinical prediction model for pregnancy (not live birth)
Ashrafi, Mahnaz, Hemat, Mandana, Arabipoor, Arezoo et al. (2017) Predictive values of anti-mullerian hormone, antral follicle count and ovarian response prediction index (ORPI) for assisted reproductive technology outcomes. Journal of obstetrics and gynaecology : the journal of the Institute of Obstetrics and Gynaecology 37(1): 82-88	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Ashrafi, Mahnaz, Rashidi, Mandana, Ghasemi, Afsaneh et al. (2013) The role of infertility etiology in success rate of intrauterine insemination cycles: an evaluation of predictive factors for pregnancy rate. International journal of fertility & sterility 7(2): 100-7	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Atasever, Melahat, Kalem, Muberra Namli, Hatirnaz, Safak et al. (2016) Factors affecting clinical pregnancy rates after IUI for the treatment of unexplained infertility and mild male subfertility. Journal of the Turkish German Gynecological Association 17(3): 134-8	- Clinical prediction model for pregnancy (not live birth)
Awadalla, Michael S, Bendikson, Kristin A, Ho, Jacqueline R et al. (2021) A validated model for predicting live birth after embryo transfer. Scientific reports 11(1): 10800	- No calibration or discrimination data can be extracted
Aydin, Yunus, Hassa, Hikmet, Oge, Tufan et al. (2013) Factors predictive of clinical pregnancy in the first intrauterine insemination cycle of 306 couples with favourable female patient characteristics. Human fertility (Cambridge, England) 16(4): 286-90	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Azantee, Y W, Murad, Z Ahmad, Roszaman, R et al. (2011) Associated factors affecting the successful pregnancy rate of intrauterine insemination at International Islamic University	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction

Study	Reason
Malaysia (IIUM) Fertility Centre. The Medical journal of Malaysia 66(3): 195-8	
Baeten, S, Bouckaert, A, Loumaye, E et al. (1993) A regression model for the rate of success of in vitro fertilization. Statistics in medicine 12(17): 1543-53	- No calibration or discrimination data can be extracted
Bahamondes, L, Alma, F A, Faundes, A et al. (1994) Score prognosis for the infertile couple based on historical factors and sperm analysis. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics 46(3): 311-5	- No calibration or discrimination data can be extracted
Baker, Valerie L, Jones, Clarence E, Cometti, Barbara et al. (2010) Factors affecting success rates in two concurrent clinical IVF trials: an examination of potential explanations for the difference in pregnancy rates between the United States and Europe. Fertility and sterility 94(4): 1287-1291	- No calibration or discrimination data can be extracted
Bakkensen, Jennifer B, Heisler, Elise, Bolten, Katherine et al. (2023) Validation of prognosis-based in vitro fertilization grant selection criteria. F&S reports 4(3): 286-291	- No calibration or discrimination data can be extracted
Ballester, Marcos, Oppenheimer, Anne, d'Argent, Emmanuelle Mathieu et al. (2012) Nomogram to predict pregnancy rate after ICSI-IVF cycle in patients with endometriosis. Human reproduction (Oxford, England) 27(2): 451-6	- Clinical prediction model for pregnancy (not live birth)
Bancsi, L F, Huijs, A M, den Ouden, C T et al. (2000) Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. Fertility and sterility 73(3): 552-7	- Clinical prediction model for pregnancy (not live birth)
Bancsi, Laszlo F J M M, Broekmans, Frank J M, Eijkemans, Marinus J C et al. (2002) Predictors of poor ovarian response in in vitro fertilization: a prospective study comparing basal markers of ovarian reserve. Fertility and sterility 77(2): 328-36	- Markers of ovarian reserve alone
Banerjee, Prajna, Choi, Bokyung, Shahine, Lora K et al. (2010) Deep phenotyping to predict live birth outcomes in in vitro fertilization.	- Data cannot be extracted

Study	Reason
Proceedings of the National Academy of Sciences of the United States of America 107(31): 13570-5	*No measure of variance reported for AUC, and no other extractable model performance statistics reported*
Bardet, L., Excoffier, J.-B., Salaun-Penquer, N. et al. (2022) Comparison of predictive models for cumulative live birth rate after treatment with ART. Reproductive BioMedicine Online 45(2): 246-255	- Data cannot be extracted *Model combines IUI and IVF/ICSI*
Barnett-Itzhaki, Zohar, Elbaz, Miriam, Butterman, Rachely et al. (2020) Machine learning vs. classic statistics for the prediction of IVF outcomes. Journal of assisted reproduction and genetics 37(10): 2405-2412	- Data cannot be extracted *No measure of variance reported for AUC, and no other extractable model performance statistics reported*
Ben-Ami, Ido, Raziel, Arieh, Strassburger, Deborah et al. (2013) Intracytoplasmic sperm injection outcome of ejaculated versus extracted testicular spermatozoa in cryptozoospermic men. Fertility and sterility 99(7): 1867-71	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Ben-Haroush, Avi, Farhi, Jacob, Zahalka, Yasmin et al. (2011) Small antral follicle count (2-5 mm) and ovarian volume for prediction of pregnancy in in vitro fertilization cycles. Gynecological endocrinology : the official journal of the International Society of Gynecological Endocrinology 27(10): 748-52	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Bensdorp, Alexandra J, van der Steeg, Jan Willem, Steures, Pieternel et al. (2017) A revised prediction model for natural conception. Reproductive biomedicine online 34(6): 619-626	- Clinical prediction model for pregnancy (not live birth)
Bi, Xuan, Feng, Long, Li, Cai et al. (2022) Modeling Pregnancy Outcomes through Sequentially Nested Regression Models. Journal of the American Statistical Association 117(538): 602-616	- Data cannot be extracted *No measure of variance reported for AUC and no other extractable model performance data reported*
Blank, Celine, Duijf, Imke T, Slappendel, Els et al. (2018) External validation of a prediction model to select the best day-three embryo for transfer in in vitro fertilization or intracytoplasmatic sperm injection procedures. Fertility and sterility 110(5): 917-924	- Model for embryo selection
Blank, Celine, Wildeboer, Rogier Rudolf, DeCroo, Ilse et al. (2019) Prediction of	- Clinical prediction model for pregnancy (not live birth)

Study	Reason
implantation after blastocyst transfer in in vitro fertilization: a machine-learning perspective. Fertility and sterility 111(2): 318-326	
Blasco, Victor, Prados, Nicolas, Carranza, Francisco et al. (2014) Influence of follicle rupture and uterine contractions on intrauterine insemination outcome: a new predictive model. Fertility and sterility 102(4): 1034-40	- No calibration or discrimination data can be extracted
Boitrelle, F, Robin, G, Marcelli, F et al. (2011) A predictive score for testicular sperm extraction quality and surgical ICSI outcome in non-obstructive azoospermia: a retrospective study. Human reproduction (Oxford, England) 26(12): 3215-21	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Bouckaert, A, Psalti, I, Loumaye, E et al. (1994) The probability of a successful treatment of infertility by in-vitro fertilization. Human reproduction (Oxford, England) 9(3): 448-55	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Brandes, M., Hamilton, C.J.C.M., Van Der Steen, J.O.M. et al. (2011) Unexplained infertility: Overall ongoing pregnancy rate and mode of conception. Human Reproduction 26(2): 360-368	- No calibration or discrimination data can be extracted
Brodin, Thomas, Hadziosmanovic, Nermin, Berglund, Lars et al. (2015) Comparing four ovarian reserve markers--associations with ovarian response and live births after assisted reproduction. Acta obstetricia et gynecologica Scandinavica 94(10): 1056-63	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Broer, Simone L, van Disseldorp, Jeroen, Broeze, Kimiko A et al. (2013) Added value of ovarian reserve testing on patient characteristics in the prediction of ovarian response and ongoing pregnancy: an individual patient data approach. Human reproduction update 19(1): 26-36	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Bu, Zhiqin, Hu, Linli, Yang, Xinhong et al. (2020) Cumulative Live Birth Rate in Patients With Thin Endometrium: A Real-World Single-Center Experience. Frontiers in endocrinology 11: 469	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Burke, L M, Davenport, A T, Russell, G B et al. (2000) Predictors of success after embryo transfer: experience from a single provider.	- No calibration or discrimination data can be extracted

Study	Reason
American journal of obstetrics and gynecology 182(5): 1001-4	
Butcher, Michael J, Janoo, Jabin, Broce, Mike et al. (2016) Use of Sperm Parameters to Predict Clinical Pregnancy with Intrauterine Insemination. The Journal of reproductive medicine 61(56): 263-9	- Clinical prediction model for pregnancy (not live birth)
Buyalos, R.P.; Daneshmand, S.; Brzechffa, P.R. (1997) Basal estradiol and follicle-stimulating hormone predict fecundity in women of advanced reproductive age undergoing ovulation induction therapy. Fertility and Sterility 68(2): 272-277	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
C N Barreto, Nayara, Castro, Giulia Z, Pereira, Ramon G et al. (2022) Predicting in vitro fertilization success in the Brazilian public health system: a machine learning approach. Medical & biological engineering & computing 60(7): 1851-1861	- Not a high-income OECD country
Cabry-Goubet, Rosalie, Scheffler, Florence, Belhadri-Mansouri, Naima et al. (2017) Effect of Gonadotropin Types and Indications on Homologous Intrauterine Insemination Success: A Study from 1251 Cycles and a Review of the Literature. BioMed research international 2017: 3512784	- No calibration or discrimination data can be extracted
Cahill, D J, Meadowcroft, J, Akande, V A et al. (2005) Likelihood of natural conception following treatment by IVF. Journal of assisted reproduction and genetics 22(1112): 401-5	- Clinical prediction model for pregnancy (not live birth)
Cai, Q F, Wan, F, Huang, R et al. (2011) Factors predicting the cumulative outcome of IVF/ICSI treatment: a multivariable analysis of 2450 patients. Human reproduction (Oxford, England) 26(9): 2532-40	- Not a high-income OECD country
Cai, Qianfang, Wan, Fei, Appleby, Dina et al. (2014) Quality of embryos transferred and progesterone levels are the most important predictors of live birth after fresh embryo transfer: a retrospective cohort study. Journal of assisted reproduction and genetics 31(2): 185-94	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction

Study	Reason
Carosso, Andrea Roberto, van Eekelen, Rik, Revelli, Alberto et al. (2022) Women in Advanced Reproductive Age: Are the Follicular Output Rate, the Follicle-Oocyte Index and the Ovarian Sensitivity Index Predictors of Live Birth in an IVF Cycle?. Journal of clinical medicine 11(3)	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Carrera-Rotllan, J; Estrada-Garcia, L; Sarquella-Ventura, J (2007) Prediction of pregnancy in IVF cycles on the fourth day of ovarian stimulation. Journal of assisted reproduction and genetics 24(9): 387-94	- Clinical prediction model for pregnancy (not live birth)
Cesarano, Sara, Pirtea, Paul, Benammar, Achraf et al. (2022) Are There Ovarian Responsive Indexes That Predict Cumulative Live Birth Rates in Women over 39 Years?. Journal of clinical medicine 11(8)	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Cheles, D.S., Dal Molin, E.A., Rocha, J.C. et al. (2020) Mining of variables from embryo morphokinetics, blastocyst's morphology and patient parameters: An approach to predict the live birth in the assisted reproduction service. Jornal Brasileiro de Reproducao Assistida 24(4): 470-479	- Non-systematic review
Chen, Hong, Sun, Zi-Li, Chen, Miao-Xin et al. (2022) Predicting the probability of a live birth after a freeze-all based in vitro fertilization-embryo transfer (IVF-ET) treatment strategy. Translational pediatrics 11(6): 797-812	- Not a high-income OECD country
Chen, Miaoxin, Huang, Xin, Liu, Yiping et al. (2020) Systematic oxidative stress is not associated with live birth rate in young non-obese patients with polycystic ovarian syndrome undergoing assisted reproduction cycles: A prospective cohort study. European journal of obstetrics, gynecology, and reproductive biology 253: 154-161	- Not a high-income OECD country
Chen, Y., Zhou, C., Ma, L. et al. (2017) Relationship between endometrial thickness and pregnancy outcome in normal responder undergoing in vitro fertilization and with transfer of two cleavage-stage embryos. Journal of Reproductive Medicine 62(5): 559-564	- Not a high-income OECD country

Study	Reason
Choi, Bokyung, Bosch, Ernesto, Lannon, Benjamin M et al. (2013) Personalized prediction of first-cycle in vitro fertilization success. Fertility and sterility 99(7): 1905-11	- Data cannot be extracted *No measure of variance reported for AUCs and no other extractable model performance data reported*
Chuang, Chih Chi, Chen, Chin-Der, Chao, Kuang-Han et al. (2003) Age is a better predictor of pregnancy potential than basal follicle-stimulating hormone levels in women undergoing in vitro fertilization. Fertility and sterility 79(1): 63-8	- Clinical prediction model for pregnancy (not live birth)
Coetzee, K, de Villiers, A, Kruger, T F et al. (1999) Clinical value of using an automated sperm morphology analyzer (IVOS). Fertility and sterility 71(2): 222-5	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Commenges-Ducos, M., Tricaud, S., Papaxanthos-Roche, A. et al. (1998) Modelling of the probability of success of the stages of in-vitro fertilization and embryo transfer: Stimulation, fertilization and implantation. Human Reproduction 13(1): 78-83	- No calibration or discrimination data can be extracted
Coppus, S F P J, van der Veen, F, Opmeer, B C et al. (2009) Evaluating prediction models in reproductive medicine. Human reproduction (Oxford, England) 24(8): 1774-8	- Non-systematic review
Corani, G., Magli, C., Giusti, A. et al. (2013) A Bayesian network model for predicting pregnancy after in vitro fertilization. Computers in Biology and Medicine 43(11): 1783-1792	- Clinical prediction model for pregnancy (not live birth)
Creus, M, Penarrubia, J, Fabregues, F et al. (2000) Day 3 serum inhibin B and FSH and age as predictors of assisted reproduction treatment outcome. Human reproduction (Oxford, England) 15(11): 2341-6	- Clinical prediction model for pregnancy (not live birth)
Custers, Inge M, Steures, Pieternel, Hompes, Peter et al. (2008) Intrauterine insemination: how many cycles should we perform?. Human reproduction (Oxford, England) 23(4): 885-8	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Custers, Inge M, Steures, Pieternel, van der Steeg, Jan Willem et al. (2007) External validation of a prediction model for an ongoing	- Clinical prediction model for pregnancy (not live birth)

Study	Reason
pregnancy after intrauterine insemination. Fertility and sterility 88(2): 425-31	
Dai, X.; Liu, J.-Y.; Wu, J. (2014) Improved prediction of in vitro fertilization clinical outcomes by combining ultrasound, serum hormones and body mass index. Journal of Reproduction and Contraception 25(3): 147-158	- Clinical prediction model for pregnancy (not live birth)
Demirdag, E., Guler, I., Akdulum, M.F.C. et al. (2022) The Impact of Serum LH levels on the Day of hCG Trigger on IVF Outcomes in Patients Undergoing GnRH Antagonist Protocols. Gazi Medical Journal 33(2): 171-174	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Deng, Y, Xie, L, Wu, XK et al. (2017) Prediction models for ovulation, conception, pregnancy and live birth in infertile women with polycystic ovary syndrome. Human reproduction (Oxford, England) 32suppl1: i454-5	- Conference abstract.
Dessolle, Lionel, Darai, Emile, Cornet, Dominique et al. (2009) Determinants of pregnancy rate in the donor oocyte model: a multivariate analysis of 450 frozen-thawed embryo transfers. Human reproduction (Oxford, England) 24(12): 3082-9	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Dessolle, Lionel, Freour, Thomas, Ravel, Celia et al. (2011) Predictive factors of healthy term birth after single blastocyst transfer. Human reproduction (Oxford, England) 26(5): 1220-6	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Duleba, A J, Hausman, N, Jones, E E et al. (1997) Preretrieval predictors of pregnancy in IVF. Journal of assisted reproduction and genetics 14(4): 205-11	- No calibration or discrimination data can be extracted
Eijkemans, Marinus J C, Imani, Babak, Mulders, Annemarie G M G J et al. (2003) High singleton live birth rate following classical ovulation induction in normogonadotrophic anovulatory infertility (WHO 2). Human reproduction (Oxford, England) 18(11): 2357-62	- Outcome is not relevant *Live birth following ovulation induction (without IUI)*
Eimers, J.M., Te Velde, E.R., Gerritse, R. et al. (1994) The prediction of the chance to conceive in subfertile couples. Fertility and Sterility 61(1): 44-52	- No calibration or discrimination data can be extracted

Study	Reason
Ejzenberg, Dani, Gomes, Tiago J O, Monteleone, Pedro A A et al. (2019) Prognostic factors for pregnancy after intrauterine insemination. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics 147(1): 65-72	- Clinical prediction model for pregnancy (not live birth)
Erdem, Ahmet, Erdem, Mehmet, Atmaca, Songul et al. (2008) Factors affecting live birth rate in intrauterine insemination cycles with recombinant gonadotrophin stimulation. Reproductive biomedicine online 17(2): 199-206	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Esteves, Sandro C, Yarali, Hakan, Vuong, Lan N et al. (2021) Cumulative delivery rate per aspiration IVF/ICSI cycle in POSEIDON patients: a real-world evidence study of 9073 patients. Human reproduction (Oxford, England) 36(8): 2157-2169	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Fang, Yuan, He, Ye, Wang, Wanlu et al. (2023) Influencing factors and predictive model of live birth involving low-grade blastocyst frozen-thawed transfer: a retrospective study. European journal of medical research 28(1): 117	- Not a high-income OECD country
Farimani, M. and Amiri, I. (2007) Analysis of prognostic factors for successful outcome in patients undergoning intrauterine insemination. Acta Medica Iranica 45(2): 101-106	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Farquhar, C M, van den Boogaard, N M, Riddell, C et al. (2011) Accessing fertility treatment in New Zealand: a comparison of the clinical priority access criteria with a prediction model for couples with unexplained subfertility. Human reproduction (Oxford, England) 26(11): 3037-44	- Data cannot be extracted *No measure of variance reported for c-statistic, and no other extractable model performance data*
Fatemi, Human M, Doody, Kevin, Griesinger, Georg et al. (2013) High ovarian response does not jeopardize ongoing pregnancy rates and increases cumulative pregnancy rates in a GnRH-antagonist protocol. Human reproduction (Oxford, England) 28(2): 442-52	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Ferlitsch, Kathrin, Sator, Michael O, Gruber, Doris M et al. (2004) Body mass index, follicle-stimulating hormone and their predictive value in in vitro fertilization. Journal of assisted reproduction and genetics 21(12): 431-6	- Clinical prediction model for pregnancy (not live birth)

Study	Reason
Franco, Jose G Jr, Petersen, Claudia G, Mauri, Ana L et al. (2017) Key performance indicators score (KPIs-score) based on clinical and laboratorial parameters can establish benchmarks for internal quality control in an ART program. JBRA assisted reproduction 21(2): 61-66	- No calibration or discrimination data can be extracted
Fu, Kaiyou, Li, Yanrui, Lv, Houyi et al. (2022) Development of a Model Predicting the Outcome of In Vitro Fertilization Cycles by a Robust Decision Tree Method. Frontiers in endocrinology 13: 877518	- Not a high-income OECD country
Fujimoto, Akihisa, Morishima, Kaoru, Harada, Miyuki et al. (2015) Elective single-embryo transfer improves cumulative pregnancy outcome in young patients but not in women of advanced reproductive age. Journal of assisted reproduction and genetics 32(12): 1773-9	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Gabbanini, Massimo, Privitera, Laura, Monzo, Ana et al. (2010) The use of prediction models of spontaneous pregnancy in in vitro fertilization units reveals differences between the expected results of public and private clinics in Spain. Fertility and sterility 94(6): 2376-8	- No calibration or discrimination data can be extracted
Gao, Hong, Liu, Dong-E, Li, Yumei et al. (2021) Early prediction of live birth for assisted reproductive technology patients: a convenient and practical prediction model. Scientific reports 11(1): 331	- Not a high-income OECD country
Gao, J., Liu, Z., Zhong, Y. et al. (2023) Factors influencing clinical pregnancy outcome of in vitro fertilization/ intracytoplasmic sperm injection in older women. African Health Sciences 23(2): 632-639	- Not a high-income OECD country
Gao, L, Li, M, Wang, Y et al. (2020) Overweight and high serum total cholesterol were risk factors for the outcome of IVF/ICSI cycles in PCOS patients and a PCOS-specific predictive model of live birth rate was established. Journal of endocrinological investigation 43(9): 1221-1228	- Not a high-income OECD country
Garrett, C., Liu, D.Y., Clarke, G.N. et al. (2003) Automated semen analysis: 'Zona pellucida preferred' sperm morphometry and straight-line	- Aim of multivariate modelling is to identify predictors associated with the outcome rather

Study	Reason
velocity are related to pregnancy rate ion subfertile couples. Human Reproduction 18(8): 1643-1649	than to develop a model for individualised prediction
Gaskins, Audrey J, Zhang, Yujia, Chang, Jeani et al. (2023) Predicted probabilities of live birth following assisted reproductive technology using United States national surveillance data from 2016 to 2018. American journal of obstetrics and gynecology 228(5): 557e1-557e10	- Data cannot be extracted *No measure of variance reported for c-statistic, and no other extractable model performance data*
Gianaroli, L, Magli, M C, Gambardella, L et al. (2013) Objective way to support embryo transfer: a probabilistic decision. Human reproduction (Oxford, England) 28(5): 1210-20	- Clinical prediction model for pregnancy (not live birth)
Goldman, Randi H, Batsis, Maria, Petrozza, John C et al. (2014) Patient-specific predictions of outcome after gonadotropin ovulation induction/intrauterine insemination. Fertility and sterility 101(6): 1649-2	- Clinical prediction model for pregnancy (not live birth)
Goldman, Randi H, Farland, Leslie V, Thomas, Ann Muir et al. (2019) The combined impact of maternal age and body mass index on cumulative live birth following in vitro fertilization. American journal of obstetrics and gynecology 221(6): 617e1-617e13	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Gong, Xiaoyun, Zhang, Yunian, Zhu, Yuejie et al. (2023) Development and validation of a live birth prediction model for expected poor ovarian response patients during IVF/ICSI. Frontiers in endocrinology 14: 1027805	- Not a high-income OECD country
Gowramma, G.S.; Nayak, S.; Cholli, N. (2021) Evaluation of machine learning algorithms on the prediction of live birth occurrence. International Journal of Pharmaceutical Research 13(2): 3243-3251	- No calibration or discrimination data can be extracted
Goyal, Ashish; Kuchana, Maheshwar; Ayyagari, Kameswari Prasada Rao (2020) Machine learning predicts live-birth occurrence before in-vitro fertilization treatment. Scientific reports 10(1): 20925	- Data cannot be extracted *No measure of variance reported for c-statistic, and no other extractable model performance data*
Gracias, R.H., Thalakottoor, L.F., Gopinath, P. et al. (2014) Fertility scoring index: CIMAR'S novel system to predict assisted reproductive	- No calibration or discrimination data can be extracted

Study	Reason
technology success. International Journal of Infertility and Fetal Medicine 5(2): 44-57	
Grzegorczyk-Martin, V., Roset, J., Di Pizio, P. et al. (2022) Adaptive data-driven models to best predict the likelihood of live birth as the IVF cycle moves on and for each embryo transfer. Journal of Assisted Reproduction and Genetics 39(8): 1937-1949	- Data cannot be extracted *No measure of variance reported for the c-statistic, and no other extractable model performance data*
Guan, Yichun, Kong, Pingping, Xiao, Zhiying et al. (2021) Independent Variables for Determining the Cumulative Live Birth Rates of Aged Patients with Polycystic Ovary Syndrome or Tubal Factor Infertility: A Retrospective Cohort Study. Frontiers in endocrinology 12: 728051	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Gunning, M.N., Christ, J.P., van Rijn, B.B. et al. (2023) Predicting pregnancy chances leading to term live birth in oligo/anovulatory women diagnosed with PCOS. Reproductive BioMedicine Online 46(1): 156-163	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Guo, Yaxin, Liu, Shuai, Hu, Shiqiao et al. (2021) High Serum Anti-Mullerian Hormone Concentrations Are Associated With Poor Pregnancy Outcome in Fresh IVF/ICSI Cycle but Not Cumulative Live Birth Rate in PCOS Patients. Frontiers in endocrinology 12: 673284	- Markers of ovarian reserve alone
Guvenir, H Altay, Misirli, Gizem, Dilbaz, Serdar et al. (2015) Estimating the chance of success in IVF treatment using a ranking algorithm. Medical & biological engineering & computing 53(9): 911-20	- Not a high-income OECD country
Guzick, D S, Balmaceda, J P, Ord, T et al. (1989) The importance of egg and sperm factors in predicting the likelihood of pregnancy from gamete intrafallopian transfer. Fertility and sterility 52(5): 795-800	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Guzman, L, Ortega-Hrepich, C, Polyzos, N P et al. (2013) A prediction model to select PCOS patients suitable for IVM treatment based on anti-Mullerian hormone and antral follicle count. Human reproduction (Oxford, England) 28(5): 1261-6	- Markers of ovarian reserve alone

Study	Reason
Haadsma, M L, Groen, H, Fidler, V et al. (2008) The predictive value of ovarian reserve tests for spontaneous pregnancy in subfertile ovulatory women. Human reproduction (Oxford, England) 23(8): 1800-7	- No calibration or discrimination data can be extracted
Hafezi, S.G., Zand, M.A., Molaei, M. et al. (2019) Dynamic model with factors of polycystic ovarian syndrome in infertile women. International Journal of Reproductive BioMedicine 17(4): 231-244	- No calibration or discrimination data can be extracted
Hafiz, Pegah, Nematollahi, Mohtaram, Boostani, Reza et al. (2017) Predicting Implantation Outcome of In Vitro Fertilization and Intracytoplasmic Sperm Injection Using Data Mining Techniques. International journal of fertility & sterility 11(3): 184-190	- Clinical prediction model for pregnancy (not live birth)
Hassan M, Al-Insaif S, Hossain M et al. (2020) A machine learning approach for prediction of pregnancy outcome following IVF treatment. Neural computing and applications 32: 2283-2297	- Not a high-income OECD country
Hazout, Andre, Bouchard, Philippe, Seifer, David B et al. (2004) Serum antimullerian hormone/mullerian-inhibiting substance appears to be a more discriminatory marker of assisted reproductive technology outcome than follicle-stimulating hormone, inhibin B, or estradiol. Fertility and sterility 82(5): 1323-9	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Hill, Micah J, Healy, Mae Wu, Richter, Kevin S et al. (2017) Revisiting the progesterone to oocyte ratio. Fertility and sterility 107(3): 671-676e2	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Hirst, W Mark, Vail, Andy, Brison, Daniel R et al. (2011) Prognostic factors influencing fresh and frozen IVF outcomes: an analysis of the UK national database. Reproductive biomedicine online 22(5): 437-48	- Data cannot be extracted *No measure of variance reported for AUC, and no other extractable model performance statistics reported*
Holschbach, V., Kordes, H., Dietrich, J.E. et al. (2023) Patient- and cycle-specific factors affecting the outcome of frozen-thawed embryo transfers. Archives of Gynecology and Obstetrics 307(6): 2001-2010	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction

Study	Reason
Homburg, Roy, Meltcer, Simion, Rabinson, Jacob et al. (2009) Do stimulation characteristics of the first in vitro fertilization cycle predict pregnancy in women of 40 years old and over?. Fertility and sterility 91(4suppl): 1311-3	- No multivariate analysis reported
Hughes, E G; King, C; Wood, E C (1989) A prospective study of prognostic factors in in vitro fertilization and embryo transfer. Fertility and sterility 51(5): 838-44	- No calibration or discrimination data can be extracted
Hunault, C C, Habbema, J D F, Eijkemans, M J C et al. (2004) Two new prediction rules for spontaneous pregnancy leading to live birth among subfertile couples, based on the synthesis of three previous models. Human reproduction (Oxford, England) 19(9): 2019-26	- Data cannot be extracted *No measure of variance reported for model performance statistics*
Hunault, Claudine C, Eijkemans, Marinus J C, Pieters, Math H E C et al. (2002) A prediction model for selecting patients undergoing in vitro fertilization for elective single embryo transfer. Fertility and sterility 77(4): 725-32	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Hunault, Claudine C, Eijkemans, Marinus J C, te Velde, Egbert R et al. (2002) Validation of a model predicting spontaneous pregnancy among subfertile untreated couples. Fertility and sterility 78(3): 500-6	- Data cannot be extracted *No measure of variance reported for model performance statistics*
Hunault, Claudine C, te Velde, Egbert R, Weima, Sjerp M et al. (2007) A case study of the applicability of a prediction model for the selection of patients undergoing in vitro fertilization for single embryo transfer in another center. Fertility and sterility 87(6): 1314-21	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Huniadi, A., Bimbo-Szuhai, E., Botea, M. et al. (2023) Fertility Predictors in Intrauterine Insemination (IUI). Journal of Personalized Medicine 13(3): 395	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Iberico, Gilberto, Vioque, Jesus, Ariza, Nuria et al. (2004) Analysis of factors influencing pregnancy rates in homologous intrauterine insemination. Fertility and sterility 81(5): 1308-13	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction

Study	Reason
Isono, Wataru, Wada-Hiraike, Osamu, Akino, Nana et al. (2019) The efficacy of non-assisted reproductive technology treatment might be limited in infertile patients with advanced endometriosis in their 30s. The journal of obstetrics and gynaecology research 45(2): 368-375	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Jenkins, J.; Hunt, L.P.; Ford, W.C.L. (2005) A model based on prognostic factors predicts the probability of success after intrauterine insemination in infertile couples. Evidence-based Obstetrics and Gynecology 7(2): 82-83	- Non-primary study (commentary)
Jia, L., Chen, P.-Y., Guo, Y.-C. et al. (2020) Prediction of cumulative live birth rate in women aged 40 years and over undergoing in vitro fertilization/intracytoplasmic sperm injectiona. Reproductive and Developmental Medicine 4(4): 233-238	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Jiang, Songwei, Li, Liuming, Li, Feiwen et al. (2020) Establishment of predictive model for analyzing clinical pregnancy outcome based on IVF-ET and ICSI assisted reproductive technology. Saudi journal of biological sciences 27(4): 1049-1056	- Not a high-income OECD country
Jiang, Xiaohua, Liu, Ruijun, Liao, Ting et al. (2021) A Predictive Model of Live Birth Based on Obesity and Metabolic Parameters in Patients With PCOS Undergoing Frozen-Thawed Embryo Transfer. Frontiers in endocrinology 12: 799871	- Not a high-income OECD country
Jin, C.; Zong, J.; Xue, S. (2022) Development of a Diagnosis Grading System for Patients Undergoing Intrauterine Inseminations: A Machine-learning Perspective. medRxiv	- Pre-publication database
Jones, Christopher A, Christensen, Anna L, Salihu, Hamisu et al. (2011) Prediction of individual probabilities of livebirth and multiple birth events following in vitro fertilization (IVF): a new outcomes counselling tool for IVF providers and patients using HFEA metrics. Journal of experimental & clinical assisted reproduction 8: 3	- Data cannot be extracted *No measure of variance reported for c-statistic, and no other extractable model performance statistics available*
Kalafat, Erkan, Benlioglu, Can, Gokce, Ali et al. (2021) Factors associated with livebirth in	- Does not include (or control for) at least 2 of the core factors (female age; duration of

Study	Reason
couples undergoing their first in vitro fertilization cycle: An internally validated prediction model. Turkish journal of obstetrics and gynecology 18(3): 212-220	subfertility; cause of subfertility; pregnancy history)
Kaleli, Semih, Kervancioglu, Mehmet E, Erol, Naciye et al. (2023) Evaluating the efficacy of ovulation stimulation with intrauterine insemination in women with diminished ovarian reserve compared to women with normal ovarian reserve. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics 160(2): 620-627	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Kato, Keiichi, Ueno, Satoshi, Yabuuchi, Akiko et al. (2014) Women's age and embryo developmental speed accurately predict clinical pregnancy after single vitrified-warmed blastocyst transfer. Reproductive biomedicine online 29(4): 411-6	- No calibration or discrimination data can be extracted
Kaufmann, S.J., Eastaugh, J.L., Snowden, S. et al. (1997) The application of neural networks in predicting the outcome of in-vitro fertilization. Human Reproduction 12(7): 1454-1457	- No calibration or discrimination data can be extracted
Kavoussi, P.K., West, B.T., Chen, S.-H. et al. (2020) A comprehensive assessment of predictors of fertility outcomes in men with non-obstructive azoospermia undergoing microdissection testicular sperm extraction. Reproductive Biology and Endocrinology 18(1): 90	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Khader, Amani, Lloyd, Suzanne M, McConnachie, Alex et al. (2013) External validation of anti-Mullerian hormone based prediction of live birth in assisted conception. Journal of ovarian research 6(1): 3	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Khafri, S., Kazemnejad, A., Movahedin, M. et al. (2008) Seasonal influences on different stages of in vitro fertilization: Stimulation and fertilization. International Journal of Fertility and Sterility 2(2): 90-95	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Khodabandelu, Sajad, Basirat, Zahra, Khaleghi, Sara et al. (2022) Developing machine learning-based models to predict intrauterine insemination (IUI) success by address modeling	- Not a high-income OECD country

Study	Reason
challenges in imbalanced data and providing modification solutions for them. BMC medical informatics and decision making 22(1): 228	
Kligman, I and Rosenwaks, Z (2001) Differentiating clinical profiles: predicting good responders, poor responders, and hyperresponders. Fertility and sterility 76(6): 1185-90	- Non-systematic review
Klipstein, Sigal, Regan, Meredith, Ryley, David A et al. (2005) One last chance for pregnancy: a review of 2,705 in vitro fertilization cycles initiated in women age 40 years and above. Fertility and sterility 84(2): 435-45	- No multivariate analysis reported
Kolte, Astrid M, Westergaard, David, Lidegaard, Ojvind et al. (2021) Chance of live birth: a nationwide, registry-based cohort study. Human reproduction (Oxford, England) 36(4): 1065-1073	- Population not relevant *Unselected population (not health-related fertility problems)*
Kozar, Nejc; Kovac, Vilma; Reljic, Milan (2021) Can methods of artificial intelligence aid in optimizing patient selection in patients undergoing intrauterine inseminations?. Journal of assisted reproduction and genetics 38(7): 1665-1673	- Clinical prediction model for pregnancy (not live birth)
La Marca, A, Nelson, S M, Sighinolfi, G et al. (2011) Anti-Mullerian hormone-based prediction model for a live birth in assisted reproduction. Reproductive biomedicine online 22(4): 341-9	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Lacey, L., Henderson, I., Hassan, S. et al. (2021) Can preoperative parameters predict successful sperm retrieval and live birth in couples undergoing testicular sperm extraction and intracytoplasmic sperm injection for azoospermia?. Middle East Fertility Society Journal 26(1): 6	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Lebovitz, Oshrit, Haas, Jigal, Mor, Nitzan et al. (2022) Predicting IVF outcome in poor ovarian responders. BMC women's health 22(1): 395	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Lee, Tsung-Hsien, Liu, Chung-Hsien, Huang, Chun-Chia et al. (2009) Impact of female age and male infertility on ovarian reserve markers to predict outcome of assisted reproduction	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction

Study	Reason
technology cycles. Reproductive biology and endocrinology : RB&E 7: 100	
Lehert, P, Arvis, P, Avril, C et al. (2021) A large observational data study supporting the PROsPeR score classification in poor ovarian responders according to live birth outcome. Human reproduction (Oxford, England) 36(6): 1600-1610	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Lehert, Philippe, Chin, Wai, Schertz, Joan et al. (2018) Predicting live birth for poor ovarian responders: the PROsPeR concept. Reproductive biomedicine online 37(1): 43-52	- Population not relevant *Restricted to those with poor ovarian response*
Lemmens, Louise, Kos, Snjezana, Beijer, Cornelis et al. (2016) Predictive value of sperm morphology and progressively motile sperm count for pregnancy outcomes in intrauterine insemination. Fertility and sterility 105(6): 1462-8	- Clinical prediction model for pregnancy (not live birth)
Leushuis, E., Van Der Steeg, J.W., Steures, P. et al. (2014) Semen analysis and prediction of natural conception. Human Reproduction 29(7): 1360-1367	- Clinical prediction model for pregnancy (not live birth)
Leushuis, Esther, van der Steeg, Jan Willem, Steures, Pieternel et al. (2009) Prediction models in reproductive medicine: a critical appraisal. Human reproduction update 15(5): 537-52	- Systematic review (not appropriate to include in its entirety but checked for eligible studies)
Levin, D.; Jun, S.H.; Dahan, M.H. (2015) Predicting pregnancy in women undergoing in-vitro fertilization with basal serum follicle stimulating hormone levels between 10.0 and 11.9 IU/L. Journal of the Turkish German Gynecology Association 16(1): 5-10	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Li, Fei, Ye, Tian, Kong, Huijuan et al. (2021) Predictive Factors for Live Birth in Fresh In Vitro Fertilization/Intracytoplasmic Sperm Injection Treatment in Poor Ovarian Reserve Patients Classified by the POSEIDON Criteria. Frontiers in endocrinology 12: 630832	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Li, Hang Wun Raymond, Lee, Vivian Chi Yan, Lau, Estella Yee Lan et al. (2013) Role of baseline antral follicle count and anti-Mullerian hormone in prediction of cumulative live birth in	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)

Study	Reason
the first in vitro fertilisation cycle: a retrospective cohort analysis. PloS one 8(4): e61095	
Liang, Rong, An, Jian, Zheng, Yijia et al. (2021) predicting and improving the probability of live birth for women undergoing frozen-thawed embryo transfer: a data-driven estimation and simulation model. Computer methods and programs in biomedicine 198: 105780	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Liao, ShuJie, Pan, Wei, Dai, Wan-Qiang et al. (2020) Development of a Dynamic Diagnosis Grading System for Infertility Using Machine Learning. JAMA network open 3(11): e2023654	- No calibration or discrimination data can be extracted
Lin, J. and Sun, X.-X. (2018) Predictive modeling in reproductive medicine. Reproductive and Developmental Medicine 2(4): 224-229	- Systematic review (not appropriate to include in its entirety but checked for eligible studies)
Lintsen, A M E, Eijkemans, M J C, Hunault, C C et al. (2007) Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study. Human reproduction (Oxford, England) 22(9): 2455-62	- Clinical prediction model for pregnancy (not live birth)
Liu, Hang, Zhang, Zhuoran, Gu, Yifan et al. (2023) Development and evaluation of a live birth prediction model for evaluating human blastocysts from a retrospective study. eLife 12	- Model for embryo selection
Liu, Ran, Bai, Shun, Jiang, Xiaohua et al. (2021) Multifactor Prediction of Embryo Transfer Outcomes Based on a Machine Learning Algorithm. Frontiers in endocrinology 12: 745039	- Clinical prediction model for pregnancy (not live birth)
Liu, Xiaoyan; Chen, Zhiyun; Ji, Yanqin (2023) Construction of the machine learning-based live birth prediction models for the first in vitro fertilization pregnant women. BMC pregnancy and childbirth 23(1): 476	- Not a high-income OECD country
Liu, Xitong, Bai, Haiyan, Shi, Wenhao et al. (2019) Frozen-thawed embryo transfer is better than fresh embryo transfer in GnRH antagonist cycle in women with 3-10 oocytes retrieved: a retrospective cohort study. Archives of gynecology and obstetrics 300(6): 1791-1796	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction

Study	Reason
Luke, Barbara, Brown, Morton B, Stern, Judy E et al. (2014) Using the Society for Assisted Reproductive Technology Clinic Outcome System morphological measures to predict live birth after assisted reproductive technology. Fertility and sterility 102(5): 1338-44	- No calibration or discrimination data can be extracted
Luke, Barbara, Brown, Morton B, Wantman, Ethan et al. (2014) A prediction model for live birth and multiple births within the first three cycles of assisted reproductive technology. Fertility and sterility 102(3): 744-52	- No calibration or discrimination data can be extracted
Luke, Barbara, Brown, Morton B, Wantman, Ethan et al. (2015) Application of a validated prediction model for in vitro fertilization: comparison of live birth rates and multiple birth rates with 1 embryo transferred over 2 cycles vs 2 embryos in 1 cycle. American journal of obstetrics and gynecology 212(5): 676e1-7	- No calibration or discrimination data can be extracted
Lundin, K.; Bergh, C.; Hardarson, T. (2001) Early embryo cleavage is a strong indicator of embryo quality in human IVF. Human Reproduction 16(12): 2652-2657	- Model for embryo selection
Luo, Yumei, Liu, Mingxing, Wu, Shunhong et al. (2022) A comprehensive evaluation of pre- and post-processing sperm parameters for predicting successful pregnancy rate following intrauterine insemination with the husband's sperms. BMC pregnancy and childbirth 22(1): 703	- Clinical prediction model for pregnancy (not live birth)
Luo, Yumei, Wu, Shunhong, Yuan, Jingru et al. (2021) Evaluation of Prognostic Factors for Clinical Pregnancy Rate Following Artificial Insemination by Husband in the Chinese Population. Frontiers in medicine 8: 638560	- Clinical prediction model for pregnancy (not live birth)
Lv, Min, Chen, Cheng, Qiu, Liping et al. (2022) A nomogram to predict extremely preterm birth in women with singleton pregnancies undergoing cervical cerclage. Heliyon 8(10): e10731	- Outcome is not relevant *Pre-term birth*
Magarelli, P C; Pearlstone, A C; Buyalos, R P (1996) Discrimination between chronological and ovarian age in infertile women aged 35 years and older: predicting pregnancy using basal follicle stimulating hormone, age and number of ovulation induction/intra-uterine	- No calibration or discrimination data can be extracted

Study	Reason
insemination cycles. Human reproduction (Oxford, England) 11(6): 1214-9	
Majumdar, G., Sengupta, A., Narad, P. et al. (2023) Deep Inception-ResNet: A Novel Approach for Personalized Prediction of Cumulative Pregnancy Outcomes in Vitro Fertilization Treatment (IVF). Journal of Obstetrics and Gynecology of India	- Clinical prediction model for pregnancy (not live birth)
Makkar, G, Ng, E H Y, Yeung, W S B et al. (2003) Prognostic factors for successful outcome in patients undergoing controlled ovarian stimulation and intrauterine insemination. Hong Kong medical journal = Xianggang yi xue za zhi 9(5): 341-5	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Malhotra, Neena, Gupta, Monica, Yadav, Anshu et al. (2021) Multivariate analysis of oocyte donor and recipient factors affecting cumulative live birth rate in oocyte donor IVF (OD-IVF) cycles. JBRA assisted reproduction 25(4): 549-556	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Malizia, Beth A; Hacker, Michele R; Penzias, Alan S (2009) Cumulative live-birth rates after in vitro fertilization. The New England journal of medicine 360(3): 236-43	- No multivariate analysis reported
Manvelyan, Evelina, Abittan, Baruch, Shan, Weiwei et al. (2023) Socioeconomic disparities in fertility treatments and associated likelihood of livebirth following in vitro fertilization. Archives of gynecology and obstetrics 308(1): 265-271	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
McDonnell, J., Goverde, A.J., Rutten, F.F.H. et al. (2002) Multivariate Markov chain analysis of the probability of pregnancy in infertile couples undergoing assisted reproduction. Human Reproduction 17(1): 103-106	- No calibration or discrimination data can be extracted
McLernon, D J, Lee, A J, Maheshwari, A et al. (2019) Predicting the chances of having a baby with or without treatment at different time points in couples with unexplained subfertility. Human reproduction (Oxford, England) 34(6): 1126-1138	- Model with or without treatment *Model performance statistics are for the ability of the model to predict live birth with or without (IVF/IUI/OS) treatment, whereas this RQ is aimed at examining performance of models that predict these separately*
McLernon, D.J. and Bhattacharya, S. (2023) Quality of clinical prediction models in in vitro	- Non-systematic review

Study	Reason
fertilisation: Which covariates are really important to predict cumulative live birth and which models are best?. Best Practice and Research: Clinical Obstetrics and Gynaecology 86: 102309	
McLernon, D.J., Steyerberg, E.W., Te Velde, E.R. et al. (2018) An improvement in the method used to assess discriminatory ability when predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation. BMJ (Online) 362: k3598	- Non-primary study (commentary)
McLernon, D.J., Te Velde, E.R., Steyerberg, E.W. et al. (2014) Clinical prediction models to inform individualized decision-making in subfertile couples: A stratified medicine approach. Human Reproduction 29(9): 1851-1858	- Non-systematic review
McLernon, David J, Raja, Edwin-Amalraj, Toner, James P et al. (2022) Predicting personalized cumulative live birth following in vitro fertilization. Fertility and sterility 117(2): 326-338	- Data cannot be extracted *No measure of variance reported for c-statistic, and no other extractable model performance statistics available*
Mehrjerd, Ameneh, Rezaei, Hassan, Eslami, Saeid et al. (2022) Internal validation and comparison of predictive models to determine success rate of infertility treatments: a retrospective study of 2485 cycles. Scientific reports 12(1): 7216	- Clinical prediction model for pregnancy (not live birth)
Meng, Shihui, Shi, Cheng, Jia, Yingying et al. (2023) A combined clinical and specific genes' model to predict live birth for in vitro fertilization and embryo transfer patients. BMC pregnancy and childbirth 23(1): 702	- Not a high-income OECD country
Merviel, Philippe, Menard, Michel, Cabry, Rosalie et al. (2021) Can Ratios Between Prognostic Factors Predict the Clinical Pregnancy Rate in an IVF/ICSI Program with a GnRH Agonist-FSH/hMG Protocol? An Assessment of 2421 Embryo Transfers, and a Review of the Literature. Reproductive sciences (Thousand Oaks, Calif.) 28(2): 495-509	- Markers of ovarian reserve alone
Milewski R, Milewska A, Więsak T et al. (2013) Comparison of artificial neural networks and logistic regression analysis in pregnancy prediction using the in vitro fertilization	- Clinical prediction model for pregnancy (not live birth)

Study	Reason
treatment. Studies in Logic, Grammar and Rhetoric 35(1): 39-48	
Minaretzis, D, Harris, D, Alper, M M et al. (1998) Multivariate analysis of factors predictive of successful live births in in vitro fertilization (IVF) suggests strategies to improve IVF outcome. Journal of assisted reproduction and genetics 15(6): 365-71	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Miyagi, Maho, Mekaru, Keiko, Nakamura, Rie et al. (2021) Live birth outcomes from IVF treatments in younger patients with low AMH. JBRA assisted reproduction 25(3): 417-421	- Does not include (or control for) at least 2 of the core factors (female age; duration of subfertility; cause of subfertility; pregnancy history)
Montanaro Gauci, M, Kruger, T F, Coetzee, K et al. (2001) Stepwise regression analysis to study male and female factors impacting on pregnancy rate in an intrauterine insemination programme. Andrologia 33(3): 135-41	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Mu, Xin, Wang, Hui, Liu, Pei-Jun et al. (2022) The interval between insemination and ovulation predicts outcome after intrauterine insemination with donor sperm (IUI-D). International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics 156(2): 341-348	- No calibration or discrimination data can be extracted
Munir, S.S., Sultana, M., Ashraf, S. et al. (2017) Predictive factors affecting success of intrauterine insemination. Pakistan Journal of Medical and Health Sciences 11(2): 747-750	- No multivariate analysis reported
Murto, Tiina, Bjuresten, Kerstin, Landgren, Britt-Marie et al. (2013) Predictive value of hormonal parameters for live birth in women with unexplained infertility and male infertility. Reproductive biology and endocrinology : RB&E 11: 61	- No calibration or discrimination data can be extracted
Nayudu, P.L., Gook, D.A., Hepworth, G. et al. (1989) Prediction of outcome in human in vitro fertilization based on follicular and stimulation response variables. Fertility and Sterility 51(1): 117-125	- Clinical prediction model for pregnancy (not live birth)
Nelson, Scott M, Fleming, Richard, Gaudoin, Marco et al. (2015) Antimullerian hormone levels and antral follicle count as prognostic indicators	- Data cannot be extracted

Study	Reason
in a personalized prediction model of live birth. Fertility and sterility 104(2): 325-32	*No measure of variance reported for AUC, and no other extractable model performance statistics reported*
Ng, Ernest Hung Yu; Yeung, William Shu Biu; Ho, Pak Chung (2005) The significance of antral follicle count in controlled ovarian stimulation and intrauterine insemination. Journal of assisted reproduction and genetics 22(910): 323-8	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Nuojua-Huttunen, S, Tomas, C, Bloigu, R et al. (1999) Intrauterine insemination treatment in subfertility: an analysis of factors affecting outcome. Human reproduction (Oxford, England) 14(3): 698-703	- Clinical prediction model for pregnancy (not live birth)
Oppenheimer, Anne, Ballester, Marcos, Mathieu d'Argent, Emmanuelle et al. (2013) Pregnancy Rate after First Intra Cytoplasmic Sperm Injection- In Vitro Fertilisation Cycle in Patients with Endometrioma with or without Deep Infiltrating Endometriosis. International journal of fertility & sterility 7(3): 207-16	- Clinical prediction model for pregnancy (not live birth)
Ottosen, Lars D M, Kesmodel, Ulrik, Hindkjaer, Johnny et al. (2007) Pregnancy prediction models and eSET criteria for IVF patients--do we need more information?. Journal of assisted reproduction and genetics 24(1): 29-36	- Clinical prediction model for pregnancy (not live birth)
Ozgur, Kemal, Bulut, Hasan, Berkkanoglu, Murat et al. (2019) Prediction of live birth and cumulative live birth rates in freeze-all-IVF treatment of a general population. Journal of assisted reproduction and genetics 36(4): 685-696	- No calibration or discrimination data can be extracted
Pan, Y., Hao, G., Wang, Q. et al. (2020) Major Factors Affecting the Live Birth Rate After Frozen Embryo Transfer Among Young Women. Frontiers in Medicine 7: 94	- Not a high-income OECD country
Pearlstone, A C, Fournet, N, Gambone, J C et al. (1992) Ovulation induction in women age 40 and older: the importance of basal follicle-stimulating hormone level and chronological age. Fertility and sterility 58(4): 674-9	- No calibration or discrimination data can be extracted

Study	Reason
Penarrubia, J.; Garcia-Velasco, J.A.; Landeras, J. (2019) Prediction models in assisted reproduction: Systematic review and critical analysis. Medicina Reproductiva y Embriologia Clinica 6(23): 63-74	- Non-English language paper
Pettersson, Goran, Andersen, Anders Nyboe, Broberg, Per et al. (2010) Pre-stimulation parameters predicting live birth after IVF in the long GnRH agonist protocol. Reproductive biomedicine online 20(5): 572-81	- Data cannot be extracted *No measure of variance reported for AUC, and no other extractable model performance statistics reported*
Pinborg, A, Gaarslev, C, Hougaard, C O et al. (2011) Influence of female bodyweight on IVF outcome: a longitudinal multicentre cohort study of 487 infertile couples. Reproductive biomedicine online 23(4): 490-9	- No calibration or discrimination data can be extracted
Pinto, Fatima, Oliveira, Cristiano, Cardoso, Margarida F et al. (2009) Impact of GnRH ovarian stimulation protocols on intracytoplasmic sperm injection outcomes. Reproductive biology and endocrinology : RB&E 7: 5	- Clinical prediction model for pregnancy (not live birth)
Porcu, Geraldine, Lehert, Philippe, Colella, Carolina et al. (2013) Predicting live birth chances for women with multiple consecutive failing IVF cycles: a simple and accurate prediction for routine medical practice. Reproductive biology and endocrinology : RB&E 11: 1	- Population not relevant *Restricted to those with multiple consecutive failing IVF cycles*
Qiu, Jiahui, Li, Pingping, Dong, Meng et al. (2019) Personalized prediction of live birth prior to the first in vitro fertilization treatment: a machine learning method. Journal of translational medicine 17(1): 317	- Not a high-income OECD country
Qu, Pengfei, Chen, Lijuan, Zhao, Doudou et al. (2022) Nomogram for the cumulative live birth in women undergoing the first IVF cycle: Base on 26, 689 patients in China. Frontiers in endocrinology 13: 900829	- Not a high-income OECD country
Raef, Behnaz and Ferdousi, Reza (2019) A Review of Machine Learning Approaches in Assisted Reproductive Technologies. Acta informatica medica : AIM : journal of the Society for Medical Informatics of Bosnia & Herzegovina : casopis Drustva za medicinsku informatiku BiH 27(3): 205-211	- Systematic review (not appropriate to include in its entirety but checked for eligible studies)

Study	Reason
Rafael, Filipa, Rodrigues, Maria Dias, Bellver, Jose et al. (2023) The combined effect of BMI and age on ART outcomes. Human reproduction (Oxford, England) 38(5): 886-894	- No calibration or discrimination data can be extracted
Ranjbari, Sima, Khatibi, Toktam, Vosough Dizaji, Ahmad et al. (2021) CNFE-SE: a novel approach combining complex network-based feature engineering and stacked ensemble to predict the success of intrauterine insemination and ranking the features. BMC medical informatics and decision making 21(1): 1	- Not a high-income OECD country
Ratna, M B, Bhattacharya, S, Abdulrahim, B et al. (2020) A systematic review of the quality of clinical prediction models in in vitro fertilisation. Human reproduction (Oxford, England) 35(1): 100-116	- Systematic review (not appropriate to include in its entirety but checked for eligible studies)
Ratna, Mariam B, Bhattacharya, Siladitya, van Geloven, N et al. (2022) Predicting cumulative live birth for couples beginning their second complete cycle of in vitro fertilization treatment. Human reproduction (Oxford, England) 37(9): 2075-2086	- Population not relevant *Prediction for 2nd complete cycle*
Rausch, Mary E, Legro, Richard S, Barnhart, Huiman X et al. (2009) Predictors of pregnancy in women with polycystic ovary syndrome. The Journal of clinical endocrinology and metabolism 94(9): 3458-66	- No calibration or discrimination data can be extracted
Reljic, Milan, Knez, Jure, Kovac, Vilma et al. (2017) Endometrial injury, the quality of embryos, and blastocyst transfer are the most important prognostic factors for in vitro fertilization success after previous repeated unsuccessful attempts. Journal of assisted reproduction and genetics 34(6): 775-779	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Reljic, Milan and Lovrec, Vida Gavric (2019) Predictive Factors for Live Birth in Autologous in Vitro Fertilization Cycles in Women Aged 40 Years and Older. Zdravstveno varstvo 58(4): 173-178	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Roberts, S A, Hirst, W M, Brison, D R et al. (2010) Embryo and uterine influences on IVF outcomes: an analysis of a UK multi-centre cohort. Human reproduction (Oxford, England) 25(11): 2792-802	- Data cannot be extracted *No measure of variance reported for AUC, and no other extractable model performance statistics reported*

Study	Reason
Roberts, S.A.; Fitzgerald, C.T.; Brison, D.R. (2009) Modelling the impact of single embryo transfer in a national health service IVF programme. Human Reproduction 24(1): 122-131	- Data cannot be extracted *No measure of variance reported for AUC, and no other extractable model performance statistics reported*
Roberts, Stephen A (2007) Models for assisted conception data with embryo-specific covariates. Statistics in medicine 26(1): 156-70	- No calibration or discrimination data can be extracted
Roseboom, T J, Vermeiden, J P, Schoute, E et al. (1995) The probability of pregnancy after embryo transfer is affected by the age of the patient, cause of infertility, number of embryos transferred and the average morphology score, as revealed by multiple logistic regression analysis. Human reproduction (Oxford, England) 10(11): 3035-41	- Clinical prediction model for pregnancy (not live birth)
Sabatini, L, Zosmer, A, Hennessy, E M et al. (2008) Relevance of basal serum FSH to IVF outcome varies with patient age. Reproductive biomedicine online 17(1): 10-9	- No calibration or discrimination data can be extracted
Saha, Laxmi, Fook-Chong, Stephanie Mc, Rajesh, Hemashree et al. (2015) Use of In Vitro Fertilisation Prediction Model in an Asian Population-Experience in Singapore. Annals of the Academy of Medicine, Singapore 44(11): 524-9	- Not a high-income OECD country
Sahin, Gulnaz, Akdogan, Aysin, Aydin, Murat Hakan et al. (2021) In-Vitro Fertilization Outcome Predictors in Women With High Baseline Follicle-Stimulating Hormone Levels: Analysis of Over 1000 Cycles From A Tertiary Center. JBRA assisted reproduction 25(2): 235-241	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Shapiro, Bruce S, Daneshmand, Said T, Garner, Forest C et al. (2008) Large blastocyst diameter, early blastulation, and low preovulatory serum progesterone are dominant predictors of clinical pregnancy in fresh autologous cycles. Fertility and sterility 90(2): 302-9	- Clinical prediction model for pregnancy (not live birth)
Shen, Lei, Zhang, Yanran, Chen, Wenfeng et al. (2022) The Application of Artificial Intelligence in Predicting Embryo Transfer Outcome of Recurrent Implantation Failure. Frontiers in physiology 13: 885661	- Clinical prediction model for pregnancy (not live birth)

Study	Reason
Shi, Wenhao, Zhang, Silin, Zhao, Wanqiu et al. (2013) Factors related to clinical pregnancy after vitrified-warmed embryo transfer: a retrospective and multivariate logistic regression analysis of 2313 transfer cycles. Human reproduction (Oxford, England) 28(7): 1768-75	- Clinical prediction model for pregnancy (not live birth)
Shibahara, Hiroaki, Obara, Hiromi, Ayustawati et al. (2004) Prediction of pregnancy by intrauterine insemination using CASA estimates and strict criteria in patients with male factor infertility. International journal of andrology 27(2): 63-8	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Shrem, G., Alasmari, N.M., Balayla, J. et al. (2021) Paternal age predicts live birth in women above 40 years of age undergoing in-vitro fertilization (IVF). Clinical and Experimental Obstetrics and Gynecology 48(2): 299-306	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Si, Manfei, Jiang, Huahua, Zhao, Yue et al. (2023) Nomogram for Predicting Live Birth after the First Fresh Embryo Transfer in Patients with PCOS Undergoing IVF/ICSI Treatment with the GnRH-Ant Protocol. Diagnostics (Basel, Switzerland) 13(11)	- Not a high-income OECD country
Simeonov, Monica, Sapir, Onit, Lande, Yechezkel et al. (2020) The entire range of trigger-day endometrial thickness in fresh IVF cycles is independently correlated with live birth rate. Reproductive biomedicine online 41(2): 239-247	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Simopoulou, M., Sfakianoudis, K., Antoniou, N. et al. (2018) Making IVF more effective through the evolution of prediction models: is prognosis the missing piece of the puzzle?. Systems Biology in Reproductive Medicine 64(5): 305-323	- Non-systematic review
Siristatidis, Charalampos, Pouliakis, Abraham, Chrelias, Charalampos et al. (2011) Artificial intelligence in IVF: a need. Systems biology in reproductive medicine 57(4): 179-85	- Non-systematic review
Siristatidis, Charalampos, Vogiatzi, Paraskevi, Pouliakis, Abraham et al. (2016) Predicting IVF Outcome: A Proposed Web-based System Using Artificial Intelligence. In vivo (Athens, Greece) 30(4): 507-12	- Protocol

Study	Reason
Smeenk, J M, Stolwijk, A M, Kremer, J A et al. (2000) External validation of the templeton model for predicting success after IVF. Human reproduction (Oxford, England) 15(5): 1065-8	- Data cannot be extracted *No measure of variance reported for AUC, and no other extractable model performance statistics reported*
Song, Jingyu, Gu, Longjie, Ren, Xinling et al. (2020) Prediction model for clinical pregnancy for ICSI after surgical sperm retrieval in different types of azoospermia. Human reproduction (Oxford, England) 35(9): 1972-1982	- Clinical prediction model for pregnancy (not live birth)
Song, Zheng, Li, Wentao, O'leary, Sean et al. (2021) Can the use of diagnostic and prognostic categorisation tailor the need for assisted reproductive technology in infertile couples?. The Australian & New Zealand journal of obstetrics & gynaecology 61(2): 297-303	- No multivariate analysis reported
Soria, Macizo, Pradillo, Galvez, Garcia, Jorquera et al. (2012) Pregnancy predictors after intrauterine insemination: analysis of 3012 cycles in 1201 couples. Journal of reproduction & infertility 13(3): 158-66	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Souter, Irene, Sun, Fangbai, Zhang, Heping et al. (2022) A personalized medicine approach to ovulation induction/ovarian stimulation: development of a predictive model and online calculator from level-I evidence. Fertility and sterility 117(2): 408-418	- Clinical prediction model for pregnancy (not live birth)
Srouji, Serene S, Mark, Alice, Levine, Zalman et al. (2005) Predicting in vitro fertilization live birth using stimulation day 6 estradiol, age, and follicle-stimulating hormone. Fertility and sterility 84(3): 795-7	- No calibration or discrimination data can be extracted
Stern, Judy E, Goldman, Marlene B, Hatasaka, Harry et al. (2009) Optimizing the number of blastocyst stage embryos to transfer on day 5 or 6 in women 38 years of age and older: a Society for Assisted Reproductive Technology database study. Fertility and sterility 91(1): 157-66	- No calibration or discrimination data can be extracted
Stern, Judy E, Goldman, Marlene B, Hatasaka, Harry et al. (2009) Optimizing the number of cleavage stage embryos to transfer on day 3 in women 38 years of age and older: a Society for Assisted Reproductive Technology database study. Fertility and sterility 91(3): 767-76	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction

Study	Reason
Steures, Pieternel, van der Steeg, Jan Willem, Mol, Ben W J et al. (2004) Prediction of an ongoing pregnancy after intrauterine insemination. Fertility and sterility 82(1): 45-51	- Clinical prediction model for pregnancy (not live birth)
Stolwijk, A M, Straatman, H, Zielhuis, G A et al. (1998) External validation of prognostic models for ongoing pregnancy after in-vitro fertilization. Human reproduction (Oxford, England) 13(12): 3542-9	- Clinical prediction model for pregnancy (not live birth)
Stolwijk, A M, Zielhuis, G A, Hamilton, C J et al. (1996) Prognostic models for the probability of achieving an ongoing pregnancy after in-vitro fertilization and the importance of testing their predictive value. Human reproduction (Oxford, England) 11(10): 2298-303	- Clinical prediction model for pregnancy (not live birth)
Sun, Xingyu, Yao, Fei, Yin, Chengliang et al. (2023) Independent value of PMOI on hCG day in predicting pregnancy outcomes in IVF/ICSI cycles. Frontiers in endocrinology 14: 1086998	- Clinical prediction model for pregnancy (not live birth)
Takahashi, Toshifumi, Hasegawa, Ayumi, Igarashi, Hideki et al. (2017) Prognostic factors for patients undergoing vitrified-warmed human embryo transfer cycles: a retrospective cohort study. Human fertility (Cambridge, England) 20(2): 140-146	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Tarin, Juan J, Pascual, Eva, Garcia-Perez, Miguel A et al. (2020) A predictive model for women's assisted fecundity before starting the first IVF/ICSI treatment cycle. Journal of assisted reproduction and genetics 37(1): 171-180	- Population not relevant *Selection based on extremes of prognosis*
Tarin, Juan J, Pascual, Eva, Gomez, Raul et al. (2020) Predictors of live birth in women with a history of biochemical pregnancies after assisted reproduction treatment. European journal of obstetrics, gynecology, and reproductive biology 248: 198-203	- Population not relevant *Selection based on prognosis*
te Velde, E R, Nieboer, D, Lintsen, A M et al. (2014) Comparison of two models predicting IVF success; the effect of time trends on model performance. Human reproduction (Oxford, England) 29(1): 57-64	- Clinical prediction model for pregnancy (not live birth)

Study	Reason
Templeton, A; Morris, J K; Parslow, W (1996) Factors that affect outcome of in-vitro fertilisation treatment. Lancet (London, England) 348(9039): 1402-6	- Data cannot be extracted *Only goodness of fit reported*
Terriou, P, Sapin, C, Giorgetti, C et al. (2001) Embryo score is a better predictor of pregnancy than the number of transferred embryos or female age. Fertility and sterility 75(3): 525-31	- No calibration or discrimination data can be extracted
Thijssen, Annelies, Creemers, An, Van der Elst, Wim et al. (2017) Predictive value of different covariates influencing pregnancy rate following intrauterine insemination with homologous semen: a prospective cohort study. Reproductive biomedicine online 34(5): 463-472	- No calibration or discrimination data can be extracted
Tian, R., Zhang, J., Xu, Y. et al. (2023) Predicting Micro-TESE among Heterogeneous Nonobstructive Azoospermic Patients: The Impact on Surgical Decision and ICSI. Andrologia 2023: 4825062	- No calibration or discrimination data can be extracted *Available for SSR but not for pregnancy or live birth*
Tjon-Kon-Fat, R I, Lar, D N, Steyerberg, E W et al. (2013) Inter-clinic variation in the chances of natural conception of subfertile couples. Human reproduction (Oxford, England) 28(5): 1391-7	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Tjon-Kon-Fat, Raissa I, Tajik, Parvin, Custers, Inge M et al. (2016) Can we identify subfertile couples that benefit from immediate in vitro fertilisation over intrauterine insemination?. European journal of obstetrics, gynecology, and reproductive biology 202: 36-40	- No multivariate analysis reported
Tomlinson, M J, Amissah-Arthur, J B, Thompson, K A et al. (1996) Prognostic indicators for intrauterine insemination (IUI): statistical model for IUI success. Human reproduction (Oxford, England) 11(9): 1892-6	- No calibration or discrimination data can be extracted
Toner, J P, Mossad, H, Grow, D R et al. (1995) Value of sperm morphology assessed by strict criteria for prediction of the outcome of artificial (intrauterine) insemination. Andrologia 27(3): 143-8	- No multivariate analysis reported
Toner, J P, Philput, C B, Jones, G S et al. (1991) Basal follicle-stimulating hormone level is a better predictor of in vitro fertilization	- No multivariate analysis reported

Study	Reason
performance than age. Fertility and sterility 55(4): 784-91	
Traub, Michael L, Van Arsdale, Anne, Pal, Lubna et al. (2009) Endometrial thickness, Caucasian ethnicity, and age predict clinical pregnancy following fresh blastocyst embryo transfer: a retrospective cohort. Reproductive biology and endocrinology : RB&E 7: 33	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Ugwu, H C, Onwuzu, S W I, Agbo, J A et al. (2022) Sonographic prediction of successful embryonic implantation in in-vitro fertilization and embryo transfer cycle procedures, using a multi-parameter approach. Radiography (London, England : 1995) 28(2): 473-479	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Vaegter, Katarina Kebbon, Lakic, Tatevik Ghukasyan, Olovsson, Matts et al. (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. Fertility and sterility 107(3): 641-648e2	- Data cannot be extracted *No measure of variance reported for AUC, and no other extractable model performance statistics reported*
van der Steeg, J W, Steures, P, Eijkemans, M J C et al. (2007) Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. Human reproduction (Oxford, England) 22(2): 536-42	- Clinical prediction model for pregnancy (not live birth)
van Eekelen, R, McLernon, D J, van Wely, M et al. (2018) External validation of a dynamic prediction model for repeated predictions of natural conception over time. Human reproduction (Oxford, England) 33(12): 2268-2275	- Clinical prediction model for pregnancy (not live birth)
van Eekelen, R, van Geloven, N, van Wely, M et al. (2019) IVF for unexplained subfertility; whom should we treat?. Human reproduction (Oxford, England) 34(7): 1249-1259	- No calibration or discrimination data can be extracted
van Eekelen, R, van Geloven, N, van Wely, M et al. (2017) Constructing the crystal ball: how to get reliable prognostic information for the management of subfertile couples. Human reproduction (Oxford, England) 32(11): 2153-2158	- Non-systematic review

Study	Reason
Van Geloven, N, Broeze, K A, Bossuyt, P M M et al. (2012) Treatment should be considered a competing risk when predicting natural conception in subfertile women. Human reproduction (Oxford, England) 27(3): 889-95	- No calibration or discrimination data can be extracted
Van Geloven, N, Van der Veen, F, Bossuyt, P M M et al. (2013) Can we distinguish between infertility and subfertility when predicting natural conception in couples with an unfulfilled child wish?. Human reproduction (Oxford, England) 28(3): 658-65	- Clinical prediction model for pregnancy (not live birth)
van Loendersloot, L L, van Wely, M, Repping, S et al. (2013) Individualized decision-making in IVF: calculating the chances of pregnancy. Human reproduction (Oxford, England) 28(11): 2972-80	- Clinical prediction model for pregnancy (not live birth)
van Loendersloot, L L, van Wely, M, Repping, S et al. (2011) Templeton prediction model underestimates IVF success in an external validation. Reproductive biomedicine online 22(6): 597-602	- Clinical prediction model for pregnancy (not live birth)
van Loendersloot, Laura, Repping, S, Bossuyt, P M M et al. (2014) Prediction models in in vitro fertilization; where are we? A mini review. Journal of advanced research 5(3): 295-301	- Non-systematic review
Van Voorhis, B J, Barnett, M, Sparks, A E et al. (2001) Effect of the total motile sperm count on the efficacy and cost-effectiveness of intrauterine insemination and in vitro fertilization. Fertility and sterility 75(4): 661-8	- No calibration or discrimination data can be extracted
van Weert, Janne-Meije, Repping, Sjoerd, van der Steeg, Jan Willem et al. (2008) A prediction model for ongoing pregnancy after in vitro fertilization in couples with male subfertility. The Journal of reproductive medicine 53(4): 250-6	- Clinical prediction model for pregnancy (not live birth)
van Weert, Janne-Meije, Repping, Sjoerd, van der Steeg, Jan Willem et al. (2005) IUI in male subfertility: are we able to select the proper patients?. Reproductive biomedicine online 11(5): 624-31	- Clinical prediction model for pregnancy (not live birth)
Villani, Maria Teresa, Morini, Daria, Spaggiari, Giorgia et al. (2021) Spontaneous pregnancies	- Aim of multivariate modelling is to identify predictors associated with the outcome rather

Study	Reason
among infertile couples during assisted reproduction lockdown for COVID-19 pandemic. Andrology 9(4): 1038-1041	than to develop a model for individualised prediction
Vogiatzi, Paraskevi; Pouliakis, Abraham; Siristatidis, Charalampos (2019) An artificial neural network for the prediction of assisted reproduction outcome. Journal of assisted reproduction and genetics 36(7): 1441-1448	- No calibration or discrimination data can be extracted
Wald, M., Sparks, A.E.T., Van Voorhis, B.J. et al. (2007) Computational models for prediction of intrauterine insemination outcomes. Journal of the Turkish German Gynecology Association 8(3): 302-307	- Clinical prediction model for pregnancy (not live birth)
Wald, Moshe, Sparks, Amy E T, Sandlow, Jay et al. (2005) Computational models for prediction of IVF/ICSI outcomes with surgically retrieved spermatozoa. Reproductive biomedicine online 11(3): 325-31	- Clinical prediction model for pregnancy (not live birth)
Wang, Cheng-Wei, Kuo, Chao-Yang, Chen, Chi-Huang et al. (2022) Predicting clinical pregnancy using clinical features and machine learning algorithms in in vitro fertilization. PloS one 17(6): e0267554	- Clinical prediction model for pregnancy (not live birth)
Wang, Kexin, Guan, Yichun, Zhang, Yuchao et al. (2023) Analysis of cumulative outcomes and influencing factors of patients with discrepancies between age and AMH levels in the early follicular phase prolonged protocol. Frontiers in endocrinology 14: 1098131	- Not a high-income OECD country
Wang, N., Yin, X., Tao, Y. et al. (2022) Cumulative live birth rates over multiple complete cycles of in vitro fertilisation cycles: 10-year cohort study of 20,687 women following freeze-all strategy from one single centre. Archives of Gynecology and Obstetrics 305(1): 251-259	- No calibration or discrimination data can be extracted
Wang, Yang, Niu, Zi-Ru, Tao, Li-Yuan et al. (2021) Nomogram to predict pregnancy outcomes of emergency oocyte freeze-thaw cycles. Chinese medical journal 134(19): 2306-2315	- Not a high-income OECD country
Weghofer, A, Barad, D H, Darmon, S K et al. (2020) The ovarian sensitivity index is predictive	- Aim of multivariate modelling is to identify predictors associated with the outcome rather

Study	Reason
of live birth chances after IVF in infertile patients. Human reproduction open 2020(4): hoaa049	than to develop a model for individualised prediction
Wen, Jen-Yu, Liu, Chung-Fen, Chung, Ming-Ting et al. (2022) Artificial intelligence model to predict pregnancy and multiple pregnancy risk following in vitro fertilization-embryo transfer (IVF-ET). Taiwanese journal of obstetrics & gynecology 61(5): 837-846	- Clinical prediction model for pregnancy (not live birth)
Wen, Mingyang, Wu, Fang, Du, Jiangbo et al. (2021) Prediction of live birth probability after in vitro fertilization and intracytoplasmic sperm injection treatment: A multi-center retrospective study in Chinese population. The journal of obstetrics and gynaecology research 47(3): 1126-1133	- Not a high-income OECD country
Wheeler, C A, Cole, B F, Frishman, G N et al. (1998) Predicting probabilities of pregnancy and multiple gestation from in vitro fertilization--a new model. Obstetrics and gynecology 91(5pt1): 696-700	- No calibration or discrimination data can be extracted
Wichmann, L; Isola, J; Tuohimaa, P (1994) Prognostic variables in predicting pregnancy. A prospective follow up study of 907 couples with an infertility problem. Human reproduction (Oxford, England) 9(6): 1102-8	- No calibration or discrimination data can be extracted
Wu, F., Liu, F., Guan, Y. et al. (2019) A nomogram predicting clinical pregnancy in the first fresh embryo transfer for women undergoing in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments. Journal of Biomedical Research 33(6): 422-429	- Clinical prediction model for pregnancy (not live birth)
Wu, Yaoqiu, Yang, Rong, Lin, Haiyan et al. (2022) A Validated Model for Individualized Prediction of Live Birth in Patients With Adenomyosis Undergoing Frozen-Thawed Embryo Transfer. Frontiers in endocrinology 13: 902083	- Not a high-income OECD country
Xi, Qingsong, Yang, Qiyu, Wang, Meng et al. (2021) Individualized embryo selection strategy developed by stacking machine learning model for better in vitro fertilization outcomes: an	- Model for embryo selection

Study	Reason
application study. Reproductive biology and endocrinology : RB&E 19(1): 53	
Xu, Boyun, Liu, Chang, Qian, Lianfen et al. (2019) Statistical Modelling Outcome of In Vitro Fertilization and Intracytoplasmic Sperm Injection: A Single Centre Study. Combinatorial chemistry & high throughput screening 22(4): 225-231	- Not a high-income OECD country
Xu, Tingting, de Figueiredo Veiga, Alexis, Hammer, Karissa C et al. (2022) Informative predictors of pregnancy after first IVF cycle using eIVF practice highway electronic health records. Scientific reports 12(1): 839	- Clinical prediction model for pregnancy (not live birth)
Yalti, S, Gurbuz, B, Sezer, H et al. (2004) Effects of semen characteristics on IUI combined with mild ovarian stimulation. Archives of andrology 50(4): 239-46	- No calibration or discrimination data can be extracted
Yang, Hongya, Liu, Fang, Ma, Yuan et al. (2022) Clinical pregnancy outcomes prediction in vitro fertilization women based on random forest prediction model: A nested case-control study. Medicine 101(49): e32232	- Clinical prediction model for pregnancy (not live birth)
Yavuz, Arzu, Demirci, Oya, Sozen, Hamdullah et al. (2013) Predictive factors influencing pregnancy rates after intrauterine insemination. Iranian journal of reproductive medicine 11(3): 227-34	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Yildirim, Goncayetkin, Turkgeldi, Lale Susan, Koroglu, Nadiye et al. (2017) Predictive factors for pregnancy outcome following controlled ovarian stimulation and intrauterine insemination. JPMA. The Journal of the Pakistan Medical Association 67(3): 422-427	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Younis, Johnny S, Jadaon, Jimmy, Izhaki, Ido et al. (2010) A simple multivariate score could predict ovarian reserve, as well as pregnancy rate, in infertile women. Fertility and sterility 94(2): 655-61	- No calibration or discrimination data can be extracted
Yu, Hsi-Cheng, Rei, Wen-May, Chiou, Shu-Ti et al. (2021) Multivariate analysis of the factors associated with live births during in vitro fertilisation in Southeast Asia: a cross-sectional study of 104,015 in vitro fertilisation records in	- Not a high-income OECD country

Study	Reason
Taiwan. Journal of assisted reproduction and genetics 38(9): 2415-2423	
Yun, Bo Hyon, Kim, Gieun, Park, Seon Hee et al. (2017) In vitro fertilization outcome in women with diminished ovarian reserve. Obstetrics & gynecology science 60(1): 46-52	- Outcome is not relevant *Models of cycle cancellation*
Zanetti, B.F., Braga, D.P.A.F., Setti, A.S. et al. (2019) Predictive factors for biochemical pregnancy in intracytoplasmic sperm injection cycles. Reproductive biology 19(1): 55-60	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Zhan, Qiong, Zhao, Jing, Paziliya, Yasheng et al. (2022) Establishing a predictive model for the evaluation of fecundity. The journal of obstetrics and gynaecology research 48(4): 987-1000	- Population not relevant *Not restricted to those with fertility problems*
Zhang, Aiping, Ma, Xiaoling, Zhang, Lili et al. (2019) Pregnancy and offspring outcomes after artificial insemination with donor sperm: A retrospective analysis of 1805 treatment cycles performed in Northwest China. Medicine 98(16): e14975	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Zhang, M., Tian, H.-Q., Bu, T. et al. (2019) Development and validation of a nomogram for predicting the probability of live birth in infertile women. Reproductive and Developmental Medicine 3(2): 77-83	- Not a high-income OECD country
Zhang, Tiancheng, Wang, Xin, Wang, Zhikai et al. (2020) A Diagnostic Model to Improve the Predictability of Natural Pregnancy Potential in Patients with Oligoasthenospermia. Medical science monitor : international medical journal of experimental and clinical research 26: e922316	- Not a high-income OECD country
Zhang, Yanran, Shen, Lei, Yin, Xinghui et al. (2022) Live-Birth Prediction of Natural-Cycle In Vitro Fertilization Using 57,558 Linked Cycle Records: A Machine Learning Perspective. Frontiers in endocrinology 13: 838087	- Data cannot be extracted *No measure of variance reported for AUC, and no other extractable model performance statistics reported*
Zhang, Z, Zhu, L-L, Jiang, H-S et al. (2016) Predictors of pregnancy outcome for infertile couples attending IVF and ICSI programmes. Andrologia 48(9): 874-881	- No calibration or discrimination data can be extracted

Study	Reason
Zhou, Feng, Zhao, Fanxuan, Jin, Xiaoying et al. (2022) Factors affecting clinical outcomes after IVF-ET for infertile young patients with ovarian endometrioma: A 5-year retrospective cohort study. Medicine 101(26): e29793	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction
Zhu, Haiyan, Zhao, Chenqiong, Xiao, Peiwen et al. (2021) Predicting the Likelihood of Live Birth in Assisted Reproductive Technology According to the Number of Oocytes Retrieved and Female Age Using a Generalized Additive Model: A Retrospective Cohort Analysis of 17,948 Cycles. Frontiers in endocrinology 12: 606231	- No calibration or discrimination data can be extracted *Mean AUC reported across models but not for the clinical prediction model*
Zippl, Anna Lena, Wachter, Alfons, Rockenschaub, Patrick et al. (2022) Predicting success of intrauterine insemination using a clinically based scoring system. Archives of gynecology and obstetrics 306(5): 1777-1786	- Clinical prediction model for pregnancy (not live birth)
Zitzmann, Michael, Nordhoff, Verena, von Schonfeld, Victoria et al. (2006) Elevated follicle-stimulating hormone levels and the chances for azoospermic men to become fathers after retrieval of elongated spermatids from cryopreserved testicular tissue. Fertility and sterility 86(2): 339-47	- Aim of multivariate modelling is to identify predictors associated with the outcome rather than to develop a model for individualised prediction

Excluded economic studies for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

No economic evidence was identified for this review.

Appendix K Research recommendations – full details

Research recommendations for review question: What is the predictive performance of clinical prediction models for assessing the chances of live birth for people with health-related fertility problems using: expectant management; intrauterine insemination (IUI); IVF with or without intracytoplasmic sperm injection (ICSI)?

No research recommendations were made for this review question.