Reporting and Analysis of Trial-Based Cost-Effectiveness Evaluations in Obstetrics and Gynaecology

Mohamed El Alili¹ · Johanna M. van Dongen¹ · Judith A. F. Huirne² · Maurits W. van Tulder¹ · Judith E. Bosmans¹

Published online: 3 July 2017

© The Author(s) 2017. This article is an open access publication

Abstract

Background and Objectives The aim was to systematically review whether the reporting and analysis of trial-based cost-effectiveness evaluations in the field of obstetrics and gynaecology comply with guidelines and recommendations, and whether this has improved over time.

Data Sources and Selection Criteria A literature search was performed in MEDLINE, the NHS Economic Evaluation Database (NHS EED) and the Health Technology Assessment (HTA) database to identify trial-based cost-effectiveness evaluations in obstetrics and gynaecology published between January 1, 2000 and May 16, 2017. Studies performed in middle- and low-income countries and studies related to prevention, midwifery, and reproduction were excluded.

Data Collection and Analysis Reporting quality was assessed using the Consolidated Health Economic Evaluation Reporting Standard (CHEERS) statement (a modified version with 21 items, as we focused on trial-based cost-effectiveness evaluations) and the statistical quality was assessed using a literature-based list of criteria (8 items). Exploratory regression analyses were performed to assess the association between reporting and statistical quality scores and publication year.

Results The electronic search resulted in 5482 potentially eligible studies. Forty-five studies fulfilled the inclusion criteria, 22 in obstetrics and 23 in gynaecology. Twenty-seven (60%) studies did not adhere to 50% ($n = 10$) or more of the reporting quality items and 32 studies (71%) did not meet 50% ($n = 4$) or more of the statistical quality items. As for the statistical quality, no study used the appropriate method to assess cost differences, no advanced methods were used to deal with missing data, and clustering of data was ignored in all studies. No significant improvements over time were found in reporting or statistical quality in gynaecology, whereas in obstetrics a significant improvement in reporting and statistical quality was found over time.

Limitations The focus of this review was on trial-based cost-effectiveness evaluations in obstetrics and gynaecology, so further research is needed to explore whether results from this review are generalizable to other medical disciplines.

Conclusions and Implications of Key Findings The reporting and analysis of trial-based cost-effectiveness evaluations in gynaecology and obstetrics is generally poor. Since this can result in biased results, incorrect conclusions, and inappropriate healthcare decisions, there is an urgent need for improvement in the methods of cost-effectiveness evaluations in this field.

Electronic supplementary material The online version of this article (doi:[10.1007/s40273-017-0531-3](https://doi.org/10.1007/s40273-017-0531-3)) contains supplementary material, which is available to authorized users.

✉ Mohamed El Alili
m.elalili@vu.nl

¹ Department of Health Sciences, Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam Public Health Research Institute, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

² Department of Obstetrics and Gynaecology, VU University Medical Centre, Amsterdam, The Netherlands

Key Points for Decision Makers

The quality of the statistical analysis and reporting of trial-based cost-effectiveness evaluations in obstetrics and gynaecology is poor with only a minority of studies presenting measures of statistical uncertainty around cost-effectiveness estimates.

Exploratory analyses indicated that there have been no significant improvements over time in reporting or statistical quality in gynaecology, whereas in obstetrics a significant improvement in reporting and statistical quality was found over time.

Improvement in reporting and statistical quality of trial-based cost-effectiveness evaluations is needed to ensure reliable results and conclusions as well as efficient allocation of scarce resources in healthcare.

1 Background

To inform decisions about the allocation of scarce healthcare resources, decision makers need information on the relative efficiency of alternative healthcare interventions, which can be provided by cost-effectiveness evaluations [1]. These cost-effectiveness evaluations are increasingly being conducted alongside controlled clinical trials (i.e. so-called trial-based cost-effectiveness evaluations) [2]. Failure to adequately conduct, analyse and/or report such cost-effectiveness evaluations can lead to biased conclusions, resulting in inappropriate healthcare decision making, and thus a possible waste of scarce resources.

A growing number of cost-effectiveness evaluations in obstetrics and gynaecology are being conducted. To illustrate, a basic MEDLINE search combining search terms related to ‘obstetrics’ and ‘gynaecology’ and the MeSH term ‘cost-benefit analysis’ showed an increase in the number of published cost-effectiveness evaluations per year, from 32 in 2000 to 112 in 2015. A large share of these cost-effectiveness evaluations were conducted alongside a clinical trial. Interventions compared in these trials often concern induction of labour, hysterectomy (i.e. surgical removal of the uterus) and care arrangement (e.g. specialist nurse providing treatment vs physician providing treatment). Outcomes of these cost-effectiveness evaluations are usually expressed in clinical outcomes; for example, the number of caesarean sections or admission to intensive care. Costs associated with these

interventions usually consist of materials used and occupation of caregiver or labour/operating room. Properly conducted cost-effectiveness evaluations in obstetrics and gynaecology can help to prevent wastage of scarce resources. This is important since obstetrics/gynaecology is a major contributor to total healthcare costs. For example, in a Dutch economic analysis comparing methods of induction, the costs of this specific obstetric procedure were estimated to be €1.4 million [3].

Reviews on the reporting and statistical methodology of trial-based cost-effectiveness evaluations show that major deficiencies are generally present in the way in which such evaluations are reported [4–7] and analysed [8–10]. This led Doshi et al. [8] to conclude that the results of trial-based cost-effectiveness evaluations need to be interpreted with caution due to the poor quality of the statistical approach. The majority of these reviews, however, only evaluated reporting quality [4–7] of trial-based cost-effectiveness evaluations and the only reviews that evaluated the statistical quality [8–10] were conducted over a decade ago. In the meantime, however, guidelines and recommendations [11–14] for trial-based cost-effectiveness evaluations have been updated and more researchers have been trained in the conduct of cost-effectiveness evaluations. In the field of obstetrics and gynaecology, methodological reviews showed similar characteristics (i.e. only evaluated reporting quality) [15, 16].

1.1 Objectives

This study aimed to explore whether the quality of reporting and the statistical methods of trial-based cost-effectiveness evaluations in obstetrics and gynaecology are in accordance with the most recent guidelines and recommendations, and whether both have improved over the past 16 years.

2 Methods

This systematic review, conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [17], included trial-based cost-effectiveness evaluations in the field of obstetrics and gynaecology that were published from January 1, 2000 up to May 16, 2017. A search was conducted in MEDLINE, the National Health Service Economic Evaluation Database (NHS EED), and the Health Technology Assessment (HTA) database. The development of the earliest guidelines took place in 1996 [18], therefore the year 2000 was used as the start date to allow for implementation of the guidelines.

2.1 Search Strategy

Databases were searched with terms related to the research field (e.g. 'gynaecology', 'obstetrics' or 'pregnancy') and study design (e.g. 'cost-utility analysis', 'economic evaluation', 'cost effectiveness' or 'economic analysis') in the title, abstract, and MeSH headings or keywords. The full PubMed search is available in Appendix S1 (see electronic supplementary material [ESM]). The electronic search was supplemented by searching reference lists of relevant review articles and of the retrieved full texts. During the search, a search log was kept consisting of keywords used, searched databases and search results. Titles and abstracts of the retrieved studies were stored in an electronic database using EndNote X7.4® (Thomson Reuters, New York, NY, US).

2.2 Study Selection

Two reviewers (ME and JMvD) independently screened titles and abstracts of identified studies for eligibility. Studies were included if they reported an economic evaluation alongside a controlled trial in obstetrics or gynaecology and concerned a cost-effectiveness analysis (CEA) and/or a cost-utility analysis (CUA). Cost-benefit analyses and cost-minimization analyses were excluded since healthcare decision makers are typically interested in CEAs and CUAs, and because statistical methods may differ across these kinds of economic evaluations [1]. Both randomized and non-randomized studies were included in the review. Papers had to be published as full papers and written in English. Furthermore, this systematic review focused on therapeutic procedures (e.g. surgical treatments, induction of labour, etc.) in obstetrics and gynaecology. Therefore, studies describing interventions related to prevention and screening as well as training of healthcare staff were excluded. Moreover, studies related to reproductive medicine (i.e. fertility) were also excluded. Finally, we specifically focused on high-income countries (e.g. countries in Europe and North America) as we expected cost-effectiveness evaluations from low-/middle-income countries to systematically be of lower quality and therefore result in significantly lower scores, whereas cost-effectiveness evaluations are mostly conducted in high-income countries (i.e. 83% of the total published cost-effectiveness evaluations) [19]. Methodological issues are typically present in cost-effectiveness evaluations from low-/middle-income countries, such as scarcity and quality of the data used, trials that do not prioritize economics and absence of cost accounting systems [20], which makes it difficult to compare evaluations between high-income and low-income countries.

Full texts were retrieved when studies fulfilled the inclusion criteria or if uncertainty remained about the inclusion of a specific study. All full texts were read and checked for eligibility by two independent reviewers (ME and JMvD). To resolve disagreement between the two reviewers, a consensus procedure was used. A third reviewer (JEB) was consulted when disagreements persisted.

2.3 Data Extraction

Two reviewers (ME and JMvD) independently extracted data from the included studies using a standardized extraction form. Agreement between the reviewers was checked during a face-to-face meeting, and a consensus procedure was used involving a third reviewer (JEB) if necessary. The first part of the extraction form focused on general study characteristics (e.g. year of publication, country), healthcare delivery (i.e. primary or secondary care), medical discipline (i.e. obstetrics or gynaecology), and the design of the trial (i.e. non-randomized study [NRS] or randomized controlled trial [RCT]). The second part focused on cost-effectiveness evaluation design aspects: type of evaluation (i.e. CEA or CUA), study perspective (e.g. healthcare perspective, societal perspective), study population, follow-up period, comparator and outcome measures. The third part focused on the statistical approach of the trial-based cost-effectiveness evaluation and is described in Sect. 2.5.

2.4 Reporting Quality of Trial-Based Cost-Effectiveness Evaluations

Reporting quality was assessed using the Consolidated Health Economic Evaluation Reporting Standard (CHEERS) statement [11] that provides concrete recommendations to optimize the reporting of cost-effectiveness evaluations. Recommendations are subdivided into six main categories: (1) title and abstract, (2) introduction, (3) methods, (4) results, (5) discussion and (6) other. For a detailed description of the CHEERS statement, the reader is referred to Husereau et al. [11]. The full CHEERS statement is provided in Appendix S2 (see ESM). As the focus of this study was to evaluate trial-based cost-effectiveness evaluations, modelling-related criteria in the statement were omitted (i.e. items 15, 16 and 18). This resulted in a modified CHEERS statement with 21 items that were answered by 'yes/no'. Studies fulfilling the criteria mentioned in the items were scored 'yes' and assigned a score of 1 per correct item ('no' was scored as 0). Answers were compared between the two reviewers and disagreements were discussed until consensus was reached. An overall reporting quality score ranging from 0 to 21 was

calculated by adding up the number of items that were scored 'yes'.

2.5 Quality of the Statistical Approach of Trial-Based Cost-Effectiveness Evaluations

To evaluate the quality of the statistical approach, four quality domains were identified based on existing guidelines [12–14]. These domains, including their subdomains, are described below.

1. **Analysis of incremental costs:** This domain consisted of three sub-domains. First, we assessed whether the cost difference was presented ('yes/no'). Studies presenting cost differences were scored as handling this sub-domain appropriately (score = 1); all others as inappropriate (score = 0). Second, we assessed the method for estimating the statistical uncertainty surrounding the cost difference, while accounting for the skewed distribution of cost data. Studies using non-parametric bootstrapping or a gamma distribution in combination with multivariable regression methods were scored as handling this sub-domain appropriately (score = 1); all others as inappropriate (score = 0) [14, 21–23]. Third, trial-based cost-effectiveness evaluations are typically underpowered for economic outcomes [24]. Consequently, researchers are recommended to use estimation (i.e. confidence intervals) rather than hypothesis testing (i.e. p values) [25]. Therefore, studies presenting confidence intervals were scored as handling this sub-domain appropriately (score = 1); all others as inappropriate (score = 0). An overall domain score was calculated by adding up the studies' scores per sub-domain (1 point per correct sub-domain, maximum score = 3).
2. **Analysis of cost-effectiveness:** This category consisted of three sub-domains. First, we assessed whether the authors presented an incremental cost-effectiveness ratio (ICER) ('yes/no'). Studies presenting an ICER were scored as handling this sub-domain appropriately (score = 1); all others as inappropriate (score = 0). Second, the method for dealing with sampling uncertainty surrounding the ICER was assessed. Non-parametric bootstrapping is considered the most appropriate method and is recommended by current guidelines [12–14]. Therefore, studies using non-parametric bootstrapping were scored as handling this sub-domain appropriately (score = 1); all others as inappropriate (score = 0). Third, we assessed whether the presentation of the uncertainty surrounding the ICER was adequate. Bootstrapped cost and effect data can be plotted in a cost-effectiveness plane (CE plane), which graphically presents the uncertainty surrounding the ICER [26]. Furthermore, the joint uncertainty surrounding costs and effects can be presented in a cost-effectiveness acceptability curve (CEAC) [27]. Presentation of 95% confidence intervals around ICERs is not considered appropriate due to interpretation issues when statistical uncertainty surrounding the ICER is distributed across more than one quadrant in the CE plane [28]. Studies presenting a CE plane and a CEAC without 95% confidence intervals around ICERs were scored as handling this sub-domain appropriately (score = 1); all others as inappropriate (score = 0). An overall domain score was calculated by adding up the studies' scores per sub-domain (1 point per correct sub-domain, maximum score = 3).
3. **Handling of missing data:** Multiple imputation (MI) is currently considered the most appropriate method for dealing with missing cost data [13, 14], while maximum likelihood approaches (e.g. expectation-maximization algorithm) are also considered to result in valid estimates [13, 29]. However, this only applies when the missingness of data has a relationship with observed factors among participants, but not with unobserved factors. This is often referred to as the Missing At Random (MAR) assumption [25, 30, 31]. Therefore, studies using one of these approaches were classified as handling this domain appropriately (score = 1); all others as inappropriate (score = 0). Furthermore, studies with only a small amount of missing data (i.e. in our review we used a threshold of $\leq 5\%$) that used a complete-case analysis were also classified as handling this domain appropriately (score = 1). When $>5\%$, but $<10\%$ of data is missing, more simple imputation techniques might be preferred over MI, purely for practical reasons [32].
4. **Addressing uncertainty (sensitivity analysis):** Three types of uncertainty are inherent to trial-based cost-effectiveness evaluations: parameter uncertainty (i.e. uncertainty due to variables that might influence results, such as unit costs), methodological uncertainty (i.e. uncertainty due to the use of different methods for analysis) and subgroup uncertainty (i.e. uncertainty due to possible differences across subgroups of participants) [33, 34]. To assess the impact of these types of uncertainty on the robustness of the results, sensitivity analyses should be undertaken [25]. Studies performing at least one of the three types of sensitivity analyses were classified as handling this domain appropriately (score = 1); all others as inappropriate (score = 0).

An overall quality score of the statistical approach, ranging from 0 to 8, was calculated per study by adding up the number of overall sub-domains that were scored 'yes'.

Table 1 Summary of appropriate methods per domain

Domain	Subdomain	Appropriate method ^a
Analysis of incremental costs	Presenting cost differences	Presented cost differences
	Estimating statistical uncertainty around cost differences	Non-parametric bootstrapping or gamma distribution combined with multivariable regression methods
	Presentation of uncertainty around cost differences	Presented confidence intervals
Analysis of cost effectiveness	Presenting ICER	Presented ICER
	Dealing with sampling uncertainty	Non-parametric bootstrapping
	Presentation of uncertainty around ICER	Presented CE plane and CEAC without confidence intervals around ICER
Handling of missing data		Multiple imputation and EM algorithm
Addressing uncertainty	Parameter uncertainty	At least one of these sensitivity analyses performed
	Methodological uncertainty	
	Subgroup analysis	

CE plane cost-effectiveness plane, *CEAC* cost-effectiveness acceptability curve, *EM* expectation-maximization, *ICER* incremental cost-effectiveness ratio

^a If the appropriate method was used, a score of 1 was rewarded. All other methods resulted in a score of 0

See Table 1 for a summary of appropriate methods per domain.

2.6 Statistical Analysis

To describe the included studies' reporting and statistical quality, descriptive statistics were used. To explore whether quality improved over time, linear regression analyses were performed; one with the overall reporting quality score as dependent variable and one with the overall quality score of the statistical approach as dependent variable stratified for medical discipline (i.e. obstetrics and gynaecology). The year of publication was used as an independent variable resulting in the regression model described below. Analyses were conducted using STATA 14[®].

$$\text{Score} = \beta_0 + \beta_1 \cdot (\text{Publication year}) + \varepsilon$$

3 Results

3.1 Literature Search and Study Selection

The electronic search identified 5482 potentially eligible studies. After removing 246 duplicates, 5236 studies were screened on title and abstract. The reviewers disagreed on the inclusion of 112 (2%) studies, resulting in an inter-rater agreement of 98%. Seventy-one studies were retrieved for full-text screening. In four cases, consensus was reached by asking a third reviewer. After the full-text screening, 44 studies [35–78] were included. One study [79] was

identified through reference checking and was also included in the review (Fig. 1). This resulted in 45 studies included for review.

3.2 Study Characteristics

Study characteristics are reported in Table 2. Just over half of the studies were conducted in gynaecology (56%; $n = 23$). Most studies conducted a CEA (87%; $n = 39$), and five (11%) studies [46, 49, 52, 69, 77] conducted a CUA. One (2%) study [79] conducted both a CEA and a CUA. The hospital perspective was used in 28 (62%) studies [35–38, 40–45, 47–50, 53, 55, 59, 60, 62, 65–68, 71, 73–75, 78], followed by the healthcare perspective (22%; $n = 10$) [39, 46, 51, 52, 54, 58, 63, 69, 70, 77] and the societal perspective (12%; $n = 5$) [56, 57, 64, 76, 79]. In two (4%) studies [61, 72], the perspective was unclear. Twenty-eight (62%) studies [35, 38–41, 48, 50, 52, 54, 55, 57, 58, 61–66, 69–77, 79] were conducted alongside an RCT and 17 (41%) [36, 37, 42–47, 49, 51, 53, 56, 59, 60, 67, 68, 78] alongside an NRS. Sample sizes ranged from 35 [55] to 9996 participants [71] and the duration of follow-up ranged from 24 hours [66] to 36 months [47]. The majority of studies were conducted in Europe (66%; $n = 27$) [35, 37, 39–41, 43, 47, 50–52, 57–59, 61–65, 67–70, 74–76, 78, 79] and North America (29%; $n = 12$) [36, 38, 42, 44–46, 48, 49, 53, 56, 60, 66]. Two (4%) studies [54, 71] were conducted over multiple countries and one (2%) study [55] did not report the country where the study was conducted, but the authors' affiliation was from the Republic of Ireland.

Fig. 1 Flow chart for inclusion of studies. *CEA* cost-effectiveness analysis, *CUA* cost-utility analysis, *HTA* Health Technology Assessment database, *NHSEED* NHS Economic Evaluation Database

3.3 Reporting Quality of the Trial-Based Cost-Effectiveness Evaluations

Results of the reporting quality assessment are presented in Table 3. The overall reporting quality score (with a maximum of 21) ranged from 1 to 17 (mean 8.8; SD 4.8; median 8). Twenty-seven (60%) studies [35–39, 42–47, 49–51, 53, 55, 56, 58–62, 66–68, 72, 78] did not adhere to $\geq 50\%$ of the items (i.e. having a score ≤ 10) of the

CHEERS statement; one (2%) study [76] had a score of 17 (81% of the items were scored positively). Criteria that were often adequately described in the studies were the title ($n = 40$; 89%), the target population ($n = 30$; 67%) and the comparators ($n = 33$; 73%). Criteria that were least appropriately described were the abstract ($n = 4$; 9%), setting and location ($n = 4$; 9%) and choice of health outcomes ($n = 6$; 13%).

Table 2 Study characteristics

References	Publication year	Data collection	Geographical area	Healthcare delivery	Medical discipline	Type of EE	Perspective	Study design	Sample size (n)	Population	Follow-up	Comparison between	Outcome measures
Bernitz et al. [35]	2012	2006–2010	Norway	Secondary care	Obstetrics	CEA	Hospital	RCT	1110	Women assessed to be at low risk at spontaneous onset of labour	From the women's admission to the hospital at onset of spontaneous labour until discharge	Midwife-led birth unit vs standard obstetric unit	Proportions of caesarean sections, instrumental vaginal deliveries, complications requiring treatment in the operating room, epidural analgesia and augmentation with oxytocin
Bienstock et al. [36]	2001	1994–1996	USA	Secondary care	Obstetrics	CEA	Hospital (not reported)	NRS	260	Patients with a history of preterm labour	Not reported	Inner-city hospital house staff vs inner city managed care organization	Primary outcomes: rate of recurrent preterm delivery Secondary outcomes: rate of NICU admission, NICU length of stay and perinatal mortality
Brooten et al. [38]	2001	1992–1996	USA	Secondary care	Obstetrics	CEA	Hospital	RCT	173	Women with high-risk pregnancies	12 months	Specialist nurse care at home vs standard prenatal care	Primary outcome: maternal effects and infant effects Secondary outcome: patient satisfaction
Eddama et al. [41]	2009	2005–2006	UK	Secondary care	Obstetrics	CEA	Hospital	RCT	350	Nulliparous women with a singleton pregnancy, cephalic presentation >37 weeks ^a , gestation, requiring cervical ripening prior to induction of labour	From randomization until hospital discharge	Isosorbide mononitrate vs placebo	Elapsed time interval from hospital admission to delivery
Eddama et al. [40]	2010	2004–2008	UK	Secondary care	Obstetrics	CEA	Hospital	RCT	500	Women before 20 weeks ^a gestation with a twin pregnancy	From randomization until hospital discharge	Vaginal progesterone gel vs placebo	Number of preterm births prevented

Table 2 continued

References	Publication year	Data collection	Geographical area	Healthcare delivery	Medical discipline	Type of EE	Perspective	Study design	Sample size (n)	Population	Follow-up	Comparison between	Outcome measures
Guo et al. [48]	2011	2001–2004	Canada	Secondary care	Obstetrics	CEA	Hospital	RCT	153	Women with clinical preterm labour	Not reported	Transdermal nitroglycerine vs placebo	Primary outcome: NICU admission Secondary outcomes: gestational age at delivery, length of NICU stay
Jakovljevic et al. [51]	2008	2004–2006	Serbia and Montenegro	Secondary care	Obstetrics	CEA	Healthcare (Republic Institute for Health Insurance in Serbia)	NRS	235	Pregnant women with threatened preterm labour	From emergence of uterine contractions to delivery	Fenoterol vs ritodrine for treatment of preterm labour	Primary outcomes: length of pregnancy, prolongation of the pregnancy, and score on modified Flanagan's quality-of-life scale for chronic diseases Secondary outcomes: quality-adjusted pregnancy weeks gained, adverse drug reactions and pregnancy outcome (neonatal health)
Lain et al. [54]	2017	2004–2013	11 countries	Secondary care	Obstetrics	CEA	Healthcare	RCT	1892	Women with a singleton pregnancy with ruptured membranes between 34 and 36 weeks' gestation	Not reported	Planned immediate birth vs delayed birth	Primary outcome: neonatal sepsis Secondary outcome: respiratory distress syndrome
Liem et al. [57]	2014	2009–2012	Netherlands	Secondary care	Obstetrics	CEA	Societal	RCT	813	Women with a multiple pregnancy	6 weeks	Cervical pessary vs standard care (no pessary)	Poor perinatal and health outcomes

Table 2 continued

References	Publication year	Data collection	Geographical area	Healthcare delivery	Medical discipline	Type of EE	Perspective	Study design	Sample size (n)	Population	Follow-up	Comparison between	Outcome measures
Morrison et al. [60]	2003	2001	USA	Secondary care	Obstetrics	CEA	Hospital inferred (not reported)	NRS	60	Women with recurrent preterm labour at <32 weeks' gestation	Not reported	Continuous subcutaneous terbutaline vs standard care	Amount of terbutaline infused and associated side effects, the gestational age at delivery, and reason for birth as well as pregnancy prolongation after discharge from the sentinel recurrent preterm labour event. Maternal hospital days, route of delivery and neonatal parameters
Niinimäki et al. [61]	2009	2003–2004	Finland	Secondary care	Obstetrics	CEA	Unclear	RCT	98	Women with a diagnosed miscarriage	2 months	Medical treatment for miscarriage vs surgical treatment for miscarriage	Success rate/ uncomplicated treatment
Petrou et al. [63]	2011	2005–2006	UK	Secondary care	Obstetrics	CEA	Healthcare (NHS)	RCT	165	Pregnant women presenting as cephalic between 36 and 41 weeks' gestation, for whom induction of labour was deemed necessary	From randomization until hospital discharge	Prostaglandin gel vs prostaglandin tablets	Time prevented between induction and delivery
Petrou et al. [64]	2006	1997–2001	UK	Secondary care	Obstetrics	CEA	Societal	RCT	1200	Women with a confirmed pregnancy of <13 weeks' gestation with a diagnosis of incomplete miscarriage or missed miscarriage	8 weeks	Expectant management vs medical or surgical management	Gynaecological infection avoided

Table 2 continued

References	Publication year	Data collection	Geographical area	Healthcare delivery	Medical discipline	Type of EE	Perspective	Study design	Sample size (n)	Population	Follow-up	Comparison between	Outcome measures
Prick et al. [65]	2014	2004–2011	Netherlands	Secondary care	Obstetrics	CEA	Hospital	RCT	519	Women with acute anaemia after postpartum haemorrhage	6 weeks	Red blood cell transfusion vs non-intervention	Primary outcome: physical fatigue Secondary outcomes: remaining health-related quality of life scores, transfusion reactions and physical complications until 6 weeks postpartum
Ramsey et al. [66]	2003	1996–1997	USA	Secondary care	Obstetrics	CEA	Hospital	RCT	111	Women with an unfavourable cervix who underwent labour induction	24 hours	Misoprostol vs dinoprostone gel or dinoprostone insert	Complete dilatation within the first 24 hours of treatment
Simon et al. [71]	2006	1998–2001	33 low-, middle- and high-income countries	Secondary care	Obstetrics	CEA	Hospital	RCT	9996	Women with pre-eclampsia	From randomization until 6 weeks, discharge from hospital after delivery or death	Magnesium sulphate vs placebo	The number of cases of eclampsia prevented or death
Sjostrom et al. [72]	2016	2011–2012	Sweden	Secondary care	Obstetrics	CEA	Unclear	RCT	1068	Healthy women seeking treatment for abortion	3 weeks	Medical abortion by physician vs medical abortion by nurse-midwife	Complete abortion without need for surgical intervention
Ten Eikelder et al. [73]	2017	2012–2013	Netherlands	Secondary care	Obstetrics	CEA	Hospital	RCT	1845	Women with a viable term singleton pregnancy in cephalic presentation, intact membranes, and unfavourable cervix without previous caesarean section	Not reported	Labour induction with oral misoprostol vs labour induction with Foley catheter	Composite safety outcome and caesarean section

Table 2 continued

References	Publication year	Data collection	Geographical area	Healthcare delivery	Medical discipline	Type of EE	Perspective	Study design	Sample size (n)	Population	Follow-up	Comparison between	Outcome measures
Van Baaren et al. [75]	2013	2009–2010	Netherlands	Secondary care	Obstetrics	CEA	Hospital	RCT	819	Pregnant women at term with an unfavourable cervix	6 weeks	Induction of labour with Foley catheter vs induction of labour with prostaglandin E2 gel	Caesarean section rate (yes/no)
Van Baaren et al. [74]	2016	2009–2013	Netherlands	Secondary care	Obstetrics	CEA	Hospital	RCT	703	Women with hypertensive disorder between 34 and 37 weeks' gestation	From randomization to hospital discharge	Immediate delivery vs expectant monitoring	Composite score of adverse maternal outcomes
Vijgen et al. [76]	2010	2005–2008	Netherlands	Secondary care	Obstetrics	CEA	Societal	RCT	756	Women diagnosed with gestational hypertension or pre-eclampsia between 36 and 41 weeks' gestation	12 months	Induction of labour vs expectant monitoring	Difference in proportion of maternal complications
Walker et al. [77]	2017	2013–	UK	Secondary care	Obstetrics	CUA	Healthcare (NHS)	RCT	241	Nulliparous women aged ≥ 35 years on their expected due date, with a singleton live fetus in a cephalic presentation	1 month	Induction of labour vs expectant monitoring	QALY
Bijen et al. [79]	2011	Unclear	Netherlands	Secondary care	Gynaecology	CEA/ CUA	Societal	RCT	279	Patients with early-stage endometrial cancer	3 months	Total laparoscopic hysterectomy vs TAH	Primary outcome CEA: major complication-free rate Primary outcome CUA: QALY
Bogliolo et al. [37]	2016	2011–2014	Italy	Secondary care	Gynaecology	CEA	Hospital inferred (not reported)	NRS	104	Women who underwent robotically assisted hysterectomy and bilateral salpingo-oophorectomy	12 months for effects and 6 months for costs	Robotic single-site hysterectomy vs multiport robotic hysterectomy	Postoperative pain, intraoperative complications, and postoperative complications

Table 2 continued

References	Publication year	Data collection	Geographical area	Healthcare delivery	Medical discipline	Type of EE	Perspective	Study design	Sample size (n)	Population	Follow-up	Comparison between	Outcome measures
Dawes et al. [39]	2007	2003–2004	UK	Secondary care	Gynaecology	CEA	Healthcare (NHS)	RCT	111	Women scheduled for major abdominal or pelvic surgery for benign gynaecological disease	6 weeks	Specialist nurse care vs standard care	Primary outcome: SF-36 health survey questionnaire Secondary outcomes: complications, length of hospital stay, readmission, information on discharge, support and satisfaction of women
El Hachem et al. [42]	2016	2013–2014	USA	Secondary care	Gynaecology	CEA	Hospital	NRS	92	Women undergoing RSS or CL	Not reported	RSS vs CL	Operative time and various perioperative outcomes
El-Sayed et al. [43]	2011	2009–2010	UK	Secondary care	Gynaecology	CEA	Hospital (inferred not reported)	NRS	140	Women with acute gynaecology conditions	Not reported	Ultrasound-based model of care vs traditional model of care	Hospital length of stay
Elabbakh et al. [44]	2000	1998–1999	USA	Secondary care	Gynaecology	CEA	Hospital (inferred not reported)	NRS	80	Obese women with early-stage endometrial carcinoma	24 months	Laparoscopic-assisted VH vs total abdominal hysterectomy	Surgical outcome, hospital stay, recall of postoperative pain control, time to return to full activity and to work, and overall satisfaction among patients
Elabbakh et al. [45]	2001	1998–1999	USA	Secondary care	Gynaecology	CEA	Hospital (inferred not reported)	NRS	147	Women with early-stage endometrial carcinoma	24 months	Laparoscopic-assisted VH vs total abdominal hysterectomy	Surgical outcome, hospital stay, recall of postoperative pain control, time to return to full activity and to work, and overall satisfaction among patients
Evans [46]	2000	Unclear	USA	Secondary care	Gynaecology	CUA	Healthcare (Medicare)	NRS	100	Patients with dysfunctional uterine bleeding	12 months	Sonohysterography vs hysteroscopic evaluation	Utility value

Table 2 continued

References	Publication year	Data collection	Geographical area	Healthcare delivery	Medical discipline	Type of EE	Perspective	Study design	Sample size (n)	Population	Follow-up	Comparison between	Outcome measures
Fernandez et al. [47]	2003	1995–1997	France	Secondary care	Gynaecology	CEA	Hospital inferred (not reported)	NRS	147	Patients who had undergone one of the three surgical interventions for menorrhagia	24–36 months	Thermo-coagulation vs VH or endometrial ablation	Primary outcome: failure rate of the method for menorrhagia Secondary outcomes: satisfaction with the procedure and ongoing pain
Horowitz et al. [49]	2002	1997–1998	USA	Secondary care	Gynaecology	CUA	Hospital inferred (not reported)	NRS	Not reported	Women undergoing gynaecological and surgical procedures	Not reported	Pre-operative autologous blood donation vs no blood donation	QALY
Jack et al. [50]	2005	2001–2002	UK	Secondary care	Gynaecology	CEA	Hospital	RCT	197	Women complaining of excessive menstrual loss	12 months	Outpatient microwave endometrial ablation vs standard microwave endometrial ablation	Primary outcomes: satisfaction with treatment and acceptability of treatment Secondary outcomes: menstrual outcomes and quality of life
Kilonzo et al. [52]	2010	2003–2005	UK	Secondary care	Gynaecology	CUA	Healthcare (NHS)	RCT	314	Women complaining of heavy menstrual bleeding	12 months	Microwave endometrial ablation vs thermal balloon endometrial ablation	QALY
Kovac [53]	2000	1988–1993	USA	Secondary care	Gynaecology	CEA	Hospital inferred (not reported)	NRS	4595	Women undergoing hysterectomy	Not reported	Decision-directed hysterectomy vs nondecision-directed hysterectomy	Primary outcome: length of stay Secondary outcome: complications
Lalchandani et al. [55]	2005	1999–2001	Not reported (Ireland and UK in authors' affiliation)	Secondary care	Gynaecology	CEA	Hospital	RCT	35	Women with minimal to moderate endometriosis	12 months	Helium thermal coagulator therapy vs medical therapy using gonadotropin-releasing hormone analogues	Mean operating time

Table 2 continued

References	Publication year	Data collection	Geographical area	Healthcare delivery	Medical discipline	Type of EE	Perspective	Study design	Sample size (n)	Population	Follow-up	Comparison between	Outcome measures
Lenihan et al. [56]	2004	2001–2003	USA	Secondary care	Gynaecology	CEA	Societal (not reported)	NRS	268	Patients that have undergone a hysterectomy	Not reported	Laparoscopic-assisted VH vs TAH or total VH	Incidence of complications, time to normal activity and return to work
Lumsden et al. [58]	2000	Unclear	UK	Secondary care	Gynaecology	CEA	Healthcare (NHS)	RCT	200	Women scheduled for an abdominal hysterectomy for benign gynaecological disease	12 months	Laparoscopic-assisted hysterectomy vs abdominal hysterectomy	Conversion rate laparoscopic-assisted VH to TAH, complication rate and quality of life
Marino et al. [59]	2015	2007–2010	France	Secondary care	Gynaecology	CEA	Hospital	NRS	306	Women referred for gynaecologic oncologic indications	24 months	Robotic-assisted laparoscopy vs standard laparoscopy	Surgical outcomes
Palomba et al. [62]	2006	2001–2003	Italy	Secondary care	Gynaecology	CEA	Hospital (not reported)	RCT	80	Postmenstrual women with severe midline pelvic pain persisting for >6 months and unresponsive to common medical treatment	12 months	Laparoscopic uterine nerve ablation vs vaginal uterosacral ligament resection	Cure rate, severity of CPP and deep dyspareunia
Ralph et al. [67]	2014	2010–2012	UK	Secondary care	Gynaecology	CEA	Hospital	NRS	90	Women undergoing VH	Not reported	ERAS vs standard care (before ERAS)	Length of inpatient stay
Sarlos et al. [68]	2010	2007–2009	Switzerland	Secondary care	Gynaecology	CEA	Hospital	NRS	80	Women needing a hysterectomy	Not reported	Robotic hysterectomy	Laparoscopic hysterectomy
Sculpher et al. [69]	2004	1999–2000	UK	Secondary care	Gynaecology	CUA	Healthcare (NHS)	RCT	487/571 ^a	Women requiring a hysterectomy for reasons other than malignancy	52 weeks	Laparoscopic hysterectomy vs VH or abdominal hysterectomy	QALY
Sculpher et al. [70]	2000	1992–1994	UK	Secondary care	Gynaecology	CEA	Healthcare	RCT	160	Pre-menopausal women with dysfunctional uterine bleeding	From randomization to 2 years after intervention	Goserelin vs danazol	Differential rate of amenorrhoea

Table 2 continued

References	Publication year	Data collection	Geographical area	Healthcare delivery	Medical discipline	Type of EE	Perspective	Study design	Sample size (n)	Population	Follow-up	Comparison between	Outcome measures
Yoong et al. [78]	2016	2009–2014	UK	Secondary care	Gynaecology	CEA	Hospital	NRS	50	Women undergoing primary vaginal or laparoscopic ovarian cystectomy for benign ovarian cysts	Not reported	Primary vaginal ovarian cystectomy vs laparoscopic approach	Patient-related outcomes

CEA cost-effectiveness analysis, CL conventional laparoscopic surgery, CPP chronic pelvic pain, CUA cost-utility analysis, ERAS enhanced recovery after surgery programme, NICU neonatal intensive care unit, NRS non-randomized study, QALY quality-adjusted life-years, RCT randomized controlled trial, RSS robotic-single-site surgery, TAH total abdominal hysterectomy, VH vaginal hysterectomy

^a Two parallel RCTs

3.4 Quality of the Statistical Approach of Trial-Based Cost-Effectiveness Evaluations

Results of the quality assessment of the statistical approach are presented in Table 4. The overall quality score of the statistical approach per study ranged from 0 to 6 (see Table 4 and Appendix S3 in ESM for scores per sub-domain). Six (15%) studies [36, 37, 46, 56, 60, 78] did not use any of the recommended methods (i.e. overall quality score = 0). Furthermore, 32 (71%) studies [35–40, 42–51, 53, 55, 56, 58–62, 65–68, 70, 72, 76, 78] did not adhere to $\geq 50\%$ of the statistical quality items (i.e. having a score ≤ 4). None of the studies (see appendix S3, ESM) used the recommended statistical method to assess the cost differences between interventions. Furthermore, no study used more advanced methods for handling missing data (i.e. multiple imputation or maximum likelihood approaches). When there was $<10\%$ missing data, more simple techniques were used in 16 (36%) studies [39, 45, 48, 49, 54, 55, 57–59, 62, 63, 66, 68, 73, 75]. Of note, no study looked into the clustered nature of the data by using methods that correct for clustering.

3.5 Improvement in Quality Over Time

Exploratory analyses showed that the reporting and statistical quality score of studies in gynaecology did not significantly improve over time. However, the statistical quality and reporting quality scores in obstetric studies did significantly improve over time. Goodness-of-fit estimates showed that the amount of variance in quality scores explained by time was only limited (Table 5).

4 Discussion

4.1 Main Findings

The majority of cost-effectiveness evaluations in obstetrics and gynaecology do not comply with current reporting guidelines and recommendations for statistical methods in trial-based cost-effectiveness evaluations. Furthermore, exploratory analyses indicated that there have not been significant improvements over time in reporting and statistical quality of trial-based cost-effectiveness evaluations in gynaecology. In obstetrics, the quality of reporting and analysis slightly improved over time.

4.2 Interpretation of the Findings

None of the included studies fully complied with the CHEERS statement's reporting criteria [11] and the median reporting quality score of the included studies was

Table 3 Reporting quality score using the CHEERS checklist

References	Title	Abstract	Background and objectives	Target population and subgroups	Setting and location	Study perspective	Comparators	Time horizon	Discount rate	Choice of health outcomes	Measurement of effectiveness
Bernitz et al. [35]	Yes	No	Yes	No	No	Yes	Yes	No	No	No	Yes
Bienstock et al. [36]	No	No	Yes	Yes	No	No	No	No	No	No	No
Brooten et al. [38]	Yes	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes
Eddama et al. [41]	Yes	No	No	Yes	No	Yes	Yes	No	Yes	No	Yes
Eddama et al. [40]	Yes	No	No	Yes	No	Yes	Yes	No	Yes	No	Yes
Guo et al. [48]	Yes	yes	No	Yes	No	Yes	Yes	No	No	No	Yes
Jakovljevic et al. [51]	Yes	No	No	Yes	Yes	Yes	Yes	No	No	No	No
Lain et al. [54]	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	No	Yes
Liem et al. [57]	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes
Morrison et al. [60]	No	No	Yes	Yes	No	No	Yes	No	No	No	No
Ninimaki et al. [61]	Yes	No	No	Yes	No	No	Yes	No	Yes	No	Yes
Petrou et al. [63]	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes	No	Yes
Petrou et al. [64]	Yes	No	No	Yes	No	Yes	Yes	No	Yes	No	Yes
Prick et al. [65]	Yes	No	Yes	Yes	Yes	No	Yes	No	No	No	Yes
Ramsey et al. [66]	Yes	No	No	Yes	No	No	Yes	No	No	No	Yes
Simon et al. [71]	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes
Sjostrom et al. [72]	Yes	No	No	No	No	No	Yes	No	Yes	No	Yes
Ten Eikelder et al. [73]	Yes	No	No	No	No	Yes	Yes	No	Yes	No	Yes
Van Baaren et al. [75]	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Van Baaren et al. [74]	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes
Viigen et al. [76]	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes
Walker et al. [77]	Yes	No	No	No	No	Yes	Yes	No	Yes	Yes	Yes
Bijen et al. [79]	Yes	No	No	Yes	No	Yes	No	No	No	Yes	Yes
Bogliolo et al. [37]	Yes	No	No	Yes	No	No	Yes	No	No	No	No
Dawes et al. [39]	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes
El Hachem et al. [42]	Yes	No	No	Yes	No	No	Yes	No	No	No	No
El-Sayed et al. [43]	Yes	No	No	Yes	No	No	Yes	No	No	No	No
Elabbakh et al. [44]	No	No	Yes	Yes	No	No	Yes	No	No	No	No
Elabbakh et al. [45]	No	No	Yes	No	No	No	Yes	No	No	No	No
Evans [46]	Yes	No	Yes	No	No	Yes	No	Yes	Yes	Yes	No
Fernandez et al. [47]	Yes	No	Yes	Yes	No	No	No	Yes	No	No	No
Horowitz et al. [49]	Yes	No	No	No	No	No	No	No	No	Yes	No
Jack et al. [50]	Yes	No	No	Yes	No	No	Yes	Yes	No	No	Yes
Kilonzo et al. [52]	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Kovac [53]	Yes	No	No	No	No	No	No	No	No	No	No

Table 3 continued

References	Title	Abstract	Background and objectives	Target population and subgroups	Setting and location	Study perspective	Comparators	Time horizon	Discount rate	Choice of health outcomes	Measurement of effectiveness
Lalchandani et al. [55]	Yes	No	No	No	No	No	Yes	No	No	No	Yes
Lenihan et al. [56]	Yes	No	No	No	No	No	No	No	No	No	No
Lumsden et al. [58]	Yes	No	No	Yes	No	Yes	No	Yes	No	No	Yes
Marino et al. [59]	Yes	No	No	No	No	Yes	No	No	No	No	No
Palomba et al. [62]	No	No	No	Yes	No	No	Yes	Yes	No	No	Yes
Reiph et al. [67]	Yes	No	No	No	No	No	No	No	No	No	No
Sarlos et al. [68]	Yes	No	No	No	No	No	Yes	No	No	No	No
Sculpher et al. [69]	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Sculpher et al. [70]	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes
Yoong et al. [78]	Yes	No	No	Yes	No	No	No	No	No	No	No
Studies complying with reporting criteria (%)											
	89	9	36	67	9	51	73	20	40	13	62
References	Measurement and valuation of preference-based outcomes	Estimating resources and costs	Currency, price date and conversion	Analytical methods	Incremental costs and outcomes	Characterizing uncertainty	Characterizing heterogeneity	Study findings, limitations, generalizability and current knowledge	Source of funding	Conflicts of interests	Score on CHEERS checklist (n yes)
Bernitz et al. [35]	NA	No	No	Yes	Yes	No	No	No	Yes	Yes	9
Bienstock et al. [36]	NA	No	No	No	No	No	No	No	No	No	2
Brooten et al. [38]	NA	No	No	No	No	No	Yes	No	Yes	No	8
Eddama et al. [41]	NA	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	13
Eddama et al. [40]	NA	Yes	No	Yes	No	Yes	No	Yes	Yes	No	11
Guo et al. [48]	NA	Yes	No	No	Yes	Yes	No	No	Yes	Yes	11
Jakovljevic et al. [51]	NA	No	No	No	No	Yes	No	No	Yes	No	7
Lain et al. [54]	NA	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	14
Liem et al. [57]	NA	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	14
Morrison et al. [60]	NA	No	No	No	No	No	No	No	No	No	3
Niinimäki et al. [61]	NA	Yes	No	Yes	Yes	No	No	No	Yes	Yes	10
Petrou et al. [63]	NA	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	15
Petrou et al. [64]	NA	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	14
Prick et al. [65]	NA	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	13
Ramsey et al. [66]	NA	No	No	No	Yes	No	No	Yes	No	No	6
Simon et al. [71]	NA	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	14
Sjöström et al. [72]	NA	No	Yes	No	Yes	No	No	Yes	Yes	No	8
Ten Eikelder et al. [73]	NA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	14
Van Baaren et al. [75]	NA	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	15

Table 3 continued

References	Measurement and valuation of preference-based outcomes	Estimating resources and costs	Currency, price date and conversion	Analytical methods	Incremental costs and outcomes	Characterizing uncertainty	Characterizing heterogeneity	Study findings, limitations, generalizability and current knowledge	Source of funding	Conflicts of interests	Score on CHEERS checklist (n yes)
Van Baaren et al. [74]	NA	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	14
Vijgen et al. [76]	NA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	17
Walker et al. [77]	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	14
Bijen et al. [79]	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	13
Bogliolo et al. [37]	NA	Yes	No	No	No	No	No	No	No	Yes	5
Dawes et al. [39]	NA	Yes	No	Yes	No	Yes	No	No	No	No	10
El Hachem et al. [42]	NA	Yes	No	No	No	No	No	Yes	No	Yes	6
El-Sayed et al. [43]	NA	No	No	No	No	No	No	No	No	Yes	2
Elabbakh et al. [44]	NA	Yes	No	No	No	No	No	No	No	No	4
Elabbakh et al. [45]	NA	Yes	No	No	No	No	No	No	No	No	3
Evans [46]	No	No	No	No	No	No	No	No	No	No	6
Fernandez et al. [47]	NA	No	No	No	Yes	No	No	No	No	No	5
Horowitz et al. [49]	No	No	No	No	No	No	Yes	Yes	No	No	4
Jack et al. [50]	NA	Yes	No	Yes	No	No	No	No	Yes	Yes	9
Kilozzo et al. [52]	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	14
Kovac [53]	NA	No	No	No	No	No	No	No	Yes	No	2
Lalchandani et al. [55]	NA	No	No	Yes	No	No	No	No	No	No	4
Lenihan et al. [56]	NA	No	No	No	No	No	No	No	No	No	1
Lumsden et al. [58]	NA	No	No	No	No	No	No	Yes	Yes	No	7
Marino et al. [59]	NA	Yes	No	Yes	No	Yes	No	No	No	No	5
Palomba et al. [62]	NA	Yes	No	Yes	No	No	Yes	No	No	No	7
Relph et al. [67]	NA	No	No	No	No	No	No	No	No	Yes	2
Sarlos et al. [68]	NA	Yes	No	Yes	No	No	No	Yes	Yes	Yes	7
Sculpher et al. [69]	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes	15
Sculpher et al. [70]	NA	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	15
Yoong et al. [78]	NA	No	No	No	No	No	No	No	No	No	2
Studies complying with reporting criteria (%)	9	60	24	49	40	47	24	42	60	53	

Compliance with reporting criteria: *italic values*: $\geq 75\%$ of reporting criteria correct; *bold values*: 51–74% of reporting criteria correct; *underlined values*: 26–50% of reporting criteria correct; *bold italic values* $\leq 25\%$ of reporting criteria correct

CHEERS Consolidated Health Economic Evaluation Reporting Standard, NA not available

Table 4 Statistical approach of included studies

References	Analysis of incremental costs			Analysis of cost effectiveness		Handling missing data	Dealing with uncertainty		Overall quality score of statistical approach
	Cost difference presented	Statistical assessment of cost differences	Presentation	ICER	Method sampling uncertainty		Parameter uncertainty	Methodological uncertainty	
Bernitz et al. [35]	No	<i>T</i> test	<i>p</i> value	Yes	Not reported, non-parametric bootstrap (1000 replications) in the sensitivity analysis	Not reported	No	Yes, non-parametric bootstrap (1000 replications) in the sensitivity analysis	2
Bienstock et al. [36]	No	<i>T</i> test	<i>p</i> value	No	Not reported	Not reported	No	No	0
Brooten et al. [38]	Yes	<i>T</i> test	<i>p</i> value	No	Not reported	Not reported	No	No	2
Eddama et al. [41]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI and <i>p</i> value	Yes	Non-parametric bootstrap (1000 replications)	Not reported	Yes	No	6
Eddama et al. [40]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI and <i>p</i> value	No	Non-parametric bootstrap (1000 replications)	Not reported	Yes	No	4
Guo et al. [48]	Yes	Not reported	No presentation	No	Not reported	Complete-case analysis <5% missing data	Yes	No	1
Jakovljevic et al. [51]	No	<i>T</i> test	<i>p</i> value	Yes	<i>T</i> test	Complete-case analysis >5% missing data	Yes	No	2
Lain et al. [54]	Yes	<i>T</i> test with bootstrap (5000 replications)	95% CI	No	Non-parametric bootstrap (5000 replications)	Complete-case analysis <5% missing data	Yes	Yes	5
Liem et al. [57]	Yes	Mann–Whitney test	95% CI	Yes	Non-parametric bootstrap (1000 replications)	Complete-case analysis <5% missing data	Yes	No	5
Morrison et al. [60]	No	<i>T</i> test	<i>p</i> value	No	Not reported	Not reported	No	No	0
Niinimäki et al. [61]	Yes	Not reported	No presentation	Yes	Not reported	Not reported	No	No	2

Table 4 continued

References	Analysis of incremental costs			Analysis of cost effectiveness		Handling missing data	Dealing with uncertainty		Overall quality score of statistical approach
	Cost difference presented	Statistical assessment of cost differences	Presentation	ICER	Method sampling uncertainty		Parameter uncertainty	Methodological uncertainty	
Petrou et al. [63]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI and <i>p</i> value	Yes	Non-parametric bootstrap (1000 replications)	Complete-case analysis <5% missing data	Yes	No	No
Petrou et al. [64]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI and <i>p</i> value	Yes	Non-parametric bootstrap (1000 replications)	Lin et al. [88] method	Yes	No	No
Prick et al. [65]	No	Not reported	No presentation	Yes	Not reported	Mean imputation	Yes	No	Yes
Ramsey et al. [66]	No	Wilcoxon rank sum test	<i>p</i> value	Yes	Not reported	No missing data	No	No	No
Simon et al. [71]	Yes	<i>T</i> test with bootstrap (?) replications	95% CI	Yes	Non-parametric bootstrap (?) replications	Mean imputation	Yes	Yes	Yes
Sjostrom et al. [72]	Yes	Unclear	No presentation	Yes	Not reported	Complete-case analysis >5% missing data	No	No	No
Ten Eikelder et al. [73]	Yes	<i>T</i> test with bootstrap (?) replications	95% CI	Yes	Non-parametric bootstrap (1000 replications)	Complete-case analysis <5% missing data	Yes	Yes	Yes
Van Baaren et al. [75]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI	Yes	Non-parametric bootstrap (1000 replications)	Complete-case analysis <5% missing data	Yes	No	Yes
Van Baaren et al. [74]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI	No	Non-parametric bootstrap (1000 replications)	Change of the perspective of the analysis	Yes	No	Yes
Vijgen et al. [76]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI	No	Non-parametric bootstrap (1000 replications)	Extrapolation	Yes	Yes	Yes

Table 4 continued

References	Analysis of incremental costs			Analysis of cost effectiveness		Handling missing data	Dealing with uncertainty		Overall quality score of statistical approach
	Cost difference presented	Statistical assessment of cost differences	Presentation	ICER	Method sampling uncertainty	Presentation sampling uncertainty	Parameter uncertainty	Methodological uncertainty	
Walker et al. [77]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI	Yes	Non-parametric bootstrap (1000 replications)	CE plane and CEAC	Yes	No	6
Bijen et al. [79]	Yes	Mann–Whitney test	<i>p</i> value	Yes	Non-parametric bootstrap (5000 replications)	CE plane and CEAC	Yes	No	6
Bogliolo et al. [37]	No	Mann–Whitney test	<i>p</i> value	No	Not reported	No presentation	No	No	0
Dawes et al. [39]	Yes	Mann–Whitney test	<i>p</i> value	No	Not reported	No presentation	Yes	No	3
El Hachem et al. [42]	Yes	<i>T</i> test or Mann–Whitney test	<i>p</i> value	No	Not reported	No presentation	No	No	1
El-Sayed et al. [43]	Yes	Not reported	No presentation	No	Not reported	No presentation	No	No	1
Elabbakh et al. [44]	Yes	<i>T</i> test	<i>p</i> value	No	Not reported	No presentation	No	No	1
Elabbakh et al. [45]	Yes	<i>T</i> test	<i>p</i> value	No	Not reported	No presentation	No	No	2
Evans [46]	No	Not reported	No presentation	No	Not reported	No presentation	No	No	0
Fernandez et al. [47]	Yes	Not reported	No presentation	Yes	Not reported	No presentation	No	No	2
Horowitz et al. [49]	No	Not reported	No presentation	Yes	Not reported	No presentation	No	No	3
Jack et al. [50]	Yes	<i>T</i> test with bootstrap (? replications)	No presentation	No	Non-parametric bootstrap (? replications)	No presentation	No	No	2

Table 4 continued

References	Analysis of incremental costs			Analysis of cost effectiveness			Handling missing data	Dealing with uncertainty		Overall quality score of statistical approach	
	Cost difference presented	Statistical assessment of cost differences	Presentation	ICER	Method sampling uncertainty	Presentation sampling uncertainty		Parameter uncertainty	Methodological uncertainty		Subgroup analysis
Kilonzo et al. [52]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI	No	Non-parametric bootstrap (1000 replications)	CE plane and CEAC	Complete-case analysis >5% missing data	Yes	Yes	No	5
Kovac [53]	Yes	Not reported	No presentation	No	Not reported	No presentation	Not reported	No	No	No	1
Lalchandani et al. [55]	No	Mann–Whitney test	<i>p</i> value	No	Not reported	No presentation	No missing data	No	No	No	1
Lenihan et al. [56]	No	ANOVA (Kruskal-Wallis)	<i>p</i> value	No	Not reported	No presentation	Complete-case analysis with >5% missing data	No	No	No	0
Lumsden et al. [58]	Yes	Not reported	95% CI	No	Not reported	No presentation	Complete-case analysis <5% missing data	No	No	No	3
Marino et al. [59]	Yes	Wilcoxon rank sum test	<i>p</i> value	No	Not reported	No presentation	Complete-case analysis <5% missing data	Yes	No	No	2
Palomba et al. [62]	No	Mann–Whitney test	<i>p</i> value	No	Not reported	No presentation	Complete-case analysis <5% missing data	No	No	Yes	2
Relph et al. [67]	Yes	Mann–Whitney test	No presentation	No	Not reported	No presentation	Complete-case analysis <5% missing data	No	No	No	1
Sarlos et al. [68]	No	Mann–Whitney test	<i>p</i> value	No	Not reported	No presentation	No missing data	No	No	No	1
Sculpher et al. [69]	Yes	<i>T</i> test with bootstrap (1000 replications)	95% CI	Yes	Non-parametric bootstrap (1000 replications)	CEAC	Lin et al. [88] method	Yes	No	No	5

Table 4 continued

References	Analysis of incremental costs			Analysis of cost effectiveness			Handling missing data		Dealing with uncertainty		Overall quality score of statistical approach
	Cost difference presented	Statistical assessment of cost differences	Presentation	ICER	Method sampling uncertainty	Presentation sampling uncertainty	Complete-case analysis	Missing data and LVCF	Parameter uncertainty	Methodological uncertainty	
Sculpher et al. [70]	Yes	Wilcoxon rank sum test	<i>p</i> value	Yes	Not reported	No presentation	Complete-case analysis >5% missing data and LVCF		Yes	No	3
Yoong et al. [78]	Yes	Wilcoxon rank sum test	<i>p</i> value	Yes	Not reported	No presentation	Complete-case analysis >5% missing data and LVCF		Yes	No	3

Compliance with statistical quality criteria: *italic values*: $\geq 75\%$ of statistical quality items correct; **bold values**: 51–74% of statistical quality items correct; underlined values: 26–50% of statistical quality items correct; **bold italic values**: $\leq 25\%$ of statistical quality items correct

CE plane cost-effectiveness plane, *CEA* cost-effectiveness analysis, *CEAC* cost-effectiveness acceptability curve, *CUA* cost-utility analysis, *ICER* incremental cost-effectiveness ratio, *LVCF* last value carried forward, *NRS* non-randomized study, *RCT* randomized controlled trial

relatively low (i.e. median 8, scale 0–21). This indicates that essential reporting components were missing, which can lead to faulty conclusions by researchers and health-care decision makers. In particular, the failure to describe the setting in which the studies were performed (i.e. the place and setting in which the resource allocation decision needs to be made such as country, primary or secondary care and healthcare system) makes it difficult to assess the relevance or transferability of cost-effectiveness evaluation results [80].

None of the included studies fully complied with the statistical recommendations extracted from existing guidelines [12–14]. Various statistical pitfalls of the included studies are noteworthy. First, some studies presented an analysis based on median costs instead of mean costs, yet the median is a measure that is not easily interpretable or usable for healthcare decision makers [25, 81, 82]. Second, ICERs were only reported by less than half of the studies. Moreover, since ICERs have well known interpretation problems, reporting 95% confidence interval surrounding ICERs is not recommended [26, 28] and presentation of uncertainty using CE planes and/or CEA curves is preferred. Nonetheless, only a small number of studies adequately presented the statistical uncertainty around the ICERs. Last, one third of the included studies relied on naïve and outdated statistical techniques for dealing with missing data (e.g. mean imputation, last observation carried forward) rather than using more advanced and valid methods such as multiple imputation and maximum likelihood approaches [83, 84]. These shortcomings in the quality of the included studies may result in either under- or overestimated cost-effectiveness outcomes.

4.3 Strengths and Limitations

A strength of this review is the systematic way in which studies were included and assessed, increasing the validity of the review. Also, to the best of our knowledge, this is the first review that combined the assessment of reporting quality with a comprehensive and in-depth evaluation of the statistical methods based on up-to-date national and international recommendations. However, several limitations need to be mentioned as well. First, in order to keep this review manageable, we focused on trial-based cost-effectiveness evaluations in obstetrics and gynaecology. Further research is needed to assess whether these results are representative of trial-based cost-effectiveness evaluations in other clinical areas. Second, reviewers may have been subjective in their judgements of quality, because they were not blinded for authors, authors' affiliations and journals. However, the quality assessments were done using objective criteria [11–14] by two independent

Table 5 Results from regression analysis for statistical quality

	Reporting quality		Statistical quality	
	Gynaecology	Obstetrics	Gynaecology	Obstetrics
β	-0.063	0.49	-0.024	0.24
95% confidence interval	-0.40; 0.28	0.20; 0.78	-0.15; 0.11	0.07; 0.42
<i>p</i> value	0.70	0.002	0.71	0.01
GOF statistic (R^2)	0.007	0.39	0.007	0.29

β refers to a decrease or increase in the quality score per publication year. Quality score could range from 0 to 21 for reporting quality and from 0 to 8 for statistical quality. Publication year could range from 2000 to 2017

GOF goodness of fit

reviewers. Third, considering the large developments in the methods of trial-based cost-effectiveness evaluations, early studies may be at a disadvantage. However, reporting guidelines have been available since 1996 [18, 85] and have not changed substantially since. Nonetheless, lower statistical quality scores may be the result of a lack of concrete, up-to-date statistical recommendations [86, 87]. Last, some of the included studies lacked transparency in how they designed and conducted their trial-based cost-effectiveness evaluations (i.e. poor reporting quality). This made it difficult to extract some of the data necessary to appropriately evaluate the quality of included studies, which affected the overall quality score negatively.

4.4 Comparison with the Literature

Our study adds to existing reviews in several ways. First, the majority of the previous reviews only assessed reporting quality and only a small number of reviews [8–10], which were conducted over a decade ago, evaluated the statistical quality of the included studies. Since then, however, statistical methods have improved considerably. Moreover, compared with previously conducted reviews in obstetrics and gynaecology, we performed an in-depth evaluation of the statistical methods.

Regardless, results of this systematic review are in line with those of previously conducted reviews, which concluded that the reporting and quality of the statistical approach of trial-based cost-effectiveness evaluations are typically poor [4–7] [8, 9] [15, 16]. However, these earlier methodological reviews in the field of obstetrics and gynaecology concluded that their quality improved over the last decades. This is in contrast with our exploratory analyses, which only showed a significant quality improvement over time in obstetrics and not in gynaecology. This discrepancy may be explained by our strict assessment of quality based on the most up-to-date evidence. All in all, our review suggests that, even though various efforts have been made during the last decade to

improve the reporting and statistical quality of trial-based cost-effectiveness evaluations, there is still substantial room for improvement in the area of obstetrics and gynaecology. Further research should indicate whether this applies to other medical disciplines as well.

4.5 Implications for Further Research and Practice

Future trial-based cost-effectiveness evaluations should increase their adherence to available guidelines and recommendations to improve their credibility. Up to now, however, no criteria list of statistical quality has been available. For this review, we developed a criteria list based on current evidence, but items were not weighed in terms of their opportunity cost; that is, the risk of taking the wrong decision. For example, failure to adequately handle missing data will affect the decisions more than evaluating cost differences using a Mann–Whitney *U* test. Therefore, we urgently recommend the development of a criteria list to assess statistical quality of trial-based cost-effectiveness evaluations including a weighing system that can be used by researchers, policy makers, reviewers and journal editors. Also, none of the most frequently used statistical software packages (e.g. SPSS, STATA, SAS, R) includes easy to use scripts for performing state-of-the-art trial-based cost-effectiveness evaluations. As such, authors are encouraged to (publicly) share their ‘advanced’ trial-based cost-effectiveness evaluations scripts.

5 Conclusion

This study indicated that the reporting and statistical quality of trial-based cost-effectiveness evaluations in obstetrics and gynaecology is generally poor. Since this can result in biased results, incorrect conclusions, and inappropriate healthcare decisions, there is an urgent need for improvement in the methods of cost-effectiveness evaluations in this field.

Data Availability Statement The authors provide the readers of this article with a data extraction sheet in which information about all included studies is summarized. This file is added as electronic supplementary material.

Authors' contributions ME: study rationale and design, literature selection, data extraction, interpretation and reflection, writing the manuscript. JvD: study rationale and design, literature selection, data extraction, interpretation and reflection, reviewing the manuscript. JH: interpretation and reflection, reviewing the manuscript. MvT: study rationale and design, interpretation and reflection, reviewing the manuscript. JEB: study rationale and design, literature selection, interpretation and reflection, reviewing the manuscript.

Compliance with Ethical Standards

Disclosure of potential conflict of interests ME reports no conflict of interest. JMvD reports no conflict of interest. JEB reports no conflict of interest. JAF has received grants from ZonMw, NOW, Samsung, Celenova and Pelgrem, outside the submitted work. MWT's institution received research grants from several government research agencies and professional organizations and his travel expenses were covered by organizing professional organizations. He received honoraria for reviewing grant proposals from Swedish and Canadian governmental grant agencies. He has not received any honoraria or travel expenses from the industry.

Ethics approval and informed consent This was a systematic review of previously published data and therefore does not require ethical approval and/or informed consent.

Funding None.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Drummond MF, Sculper MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 4th ed. Oxford: Oxford University Press; 2015.
2. Glick H, Doshi JA, Sonnad SS, Polsky D. *Economic evaluation in clinical trials*. 2nd ed. Oxford: Oxford University Press; 2015.
3. Vijgen SM, Boers KE, Opmeer BC, Bijlenga D, Bekedam DJ, Bloemenkamp KW, et al. Economic analysis comparing induction of labour and expectant management for intrauterine growth restriction at term (DIGITAT trial). *Eur J Obstet Gynecol Reprod Biol*. 2013;170(2):358–63.
4. Neumann PJ, Fang CH, Cohen JT. 30 years of pharmaceutical cost-utility analyses: growth, diversity and methodological improvement. *Pharmacoeconomics*. 2009;27(10):861–72.
5. Neumann PJ, Greenberg D, Olchanski NV, Stone PW, Rosen AB. Growth and quality of the cost-utility literature, 1976–2001. *Value Health*. 2005;8(1):3–9.
6. Neumann PJ, Stone PW, Chapman RH, Sandberg EA, Bell CM. The quality of reporting in published cost-utility analyses, 1976–1997. *Ann Intern Med*. 2000;132(12):964–72.
7. Rosen AB, Greenberg D, Stone PW, Olchanski NV, Neumann PJ. Quality of abstracts of papers reporting original cost-effectiveness analyses. *Med Decis Mak*. 2005;25(4):424–8.
8. Doshi JA, Glick HA, Polsky D. Analyses of cost data in economic evaluations conducted alongside randomized controlled trials. *Value Health*. 2006;9(5):334–40.
9. Barber JA, Thompson SG. Analysis and interpretation of cost data in randomised controlled trials: review of published studies. *BMJ*. 1998;317(7167):1195–200.
10. Udvarhelyi IS, Colditz GA, Rai A, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature—are the methods being used correctly. *Ann Intern Med*. 1992;116(3):238–44.
11. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health*. 2013;16(2):e1–5.
12. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal London, United Kingdom: NICE UK; 2013 [updated 04/04/13; cited 2016 20/04/16]. <https://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf>.
13. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Diemen: Zorginstituut Nederland; 2015.
14. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II—an ISPOR Good Research Practices Task Force report. *Value Health*. 2015;18(2):161–72.
15. Subak LL, Caughey AB, Washington AE. Cost-effectiveness analyses in obstetrics and gynecology. Evaluation of methodologic quality and trends. *J Reprod Med*. 2002;47(8):631–9.
16. Vijgen SM, Opmeer BC, Mol BW. The methodological quality of economic evaluation studies in obstetrics and gynecology: a systematic review. *Am J Perinatol*. 2013;30(4):253–60.
17. Moher D, Liberati A, Tetzlaff J, Altman DG, Group Prisma. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
18. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *Br Med J*. 1996;313(7052):275–83.
19. Pitt C, Goodman C, Hanson K. Economic evaluation in global perspective: a bibliometric analysis of the recent literature. *Health Econ*. 2016;25:9–28.
20. Pitt C, Vassall A, Teerawattananon Y, Griffiths UK, Guinness L, Walker D, et al. Foreword: health economic evaluations in low- and middle-income countries: methodological issues and challenges for priority setting. *Health Econ*. 2016;25(Suppl 1):1–5.
21. Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? *BMJ*. 2000;320(7243):1197–200.
22. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. *J Health Econ*. 2005;24(3):465–88.
23. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. *J Health Serv Res Policy*. 2004;9(4):197–204.
24. Briggs A. Economic evaluation and clinical trials: size matters. *BMJ*. 2000;321(7273):1362–3.
25. van Dongen JM, van Wier MF, Tompa E, Bongers PM, van der Beek AJ, van Tulder MW, et al. Trial-based economic evaluations in occupational health: principles, methods, and recommendations. *J Occup Environ Med*. 2014;56(6):563–72.

26. Black WC. The CE plane: a graphic representation of cost-effectiveness. *Med Decis Mak.* 1990;10(3):212–4.
27. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves—facts, fallacies and frequently asked questions. *Health Econ.* 2004;13(5):405–15.
28. Briggs AH, O'Brien BJ, Blackhouse G. Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annu Rev Public Health.* 2002;23:377–401.
29. Oostenbrink JB, Al MJ. The analysis of incomplete cost data due to dropout. *Health Econ.* 2005;14(8):763–76.
30. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393.
31. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
32. Eekhout I, de Vet HC, Twisk JW, Brand JP, de Boer MR, Heymans MW. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *J Clin Epidemiol.* 2014;67(3):335–42.
33. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Econ.* 1994;3(2):95–104.
34. Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. *Pharmacoeconomics.* 2008;26(9):799–806.
35. Bernitz S, Aas E, Oian P. Economic evaluation of birth care in low-risk women. A comparison between a midwife-led birth unit and a standard obstetric unit within the same hospital in Norway. A randomised controlled trial. *Midwifery.* 2012;28(5):591–9.
36. Bienstock JL, Ural SH, Blakemore K, Pressman EK. University hospital-based prenatal care decreases the rate of preterm delivery and costs, when compared to managed care. *J Matern-Fetal Med.* 2001;10(2):127–30.
37. Bogliolo S, Ferrero S, Cassani C, Musacchi V, Zanellini F, Dominoni M, et al. Single-site versus multiport robotic hysterectomy in benign gynecologic diseases: a retrospective evaluation of surgical outcomes and cost analysis. *J Minim Invas Gynecol.* 2016;23(4):603–9.
38. Brooten D, Youngblut JM, Brown L, Finkler SA, Neff DF, Madigan E. A randomized trial of nurse specialist home care for women with high-risk pregnancies: outcomes and costs. *Am J Manag Care.* 2001;7(8):793–803.
39. Dawes HA, Docherty T, Traynor I, Gilmore DH, Jardine AG, Knill-Jones R. Specialist nurse supported discharge in gynaecology: a randomised comparison and economic evaluation. *Eur J Obstet Gynecol Reprod Biol.* 2007;130(2):262–70.
40. Eddama O, Petrou S, Regier D, Norrie J, MacLennan G, Mackenzie F, et al. Study of progesterone for the prevention of preterm birth in twins (STOPPIT): findings from a trial-based cost-effectiveness analysis. *Int J Technol Assess Health Care.* 2010;26(2):141–8.
41. Eddama O, Petrou S, Schroeder L, Bollapragada SS, Mackenzie F, Norrie J, et al. The cost-effectiveness of outpatient (at home) cervical ripening with isosorbide mononitrate prior to induction of labour. *BJOG Int J Obstet Gynaecol.* 2009;116(9):1196–203.
42. El Hachem L, Andikyan V, Mathews S, Friedman K, Poeran J, Shieh K, et al. Robotic single-site and conventional laparoscopic surgery in gynecology: clinical outcomes and cost analysis of a matched case-control study. *J Minim Invas Gynecol.* 2016;23(5):760–8.
43. El-Sayed MM, Mohamed SA, Jones MH. Cost-effectiveness of ultrasound use by on-call registrars in an acute gynaecology setting. *J Obstet Gynaecol.* 2011;31(8):743–5.
44. Eltabbakh GH, Shamonki MI, Moody JM, Garafano LL. Hysterectomy for obese women with endometrial cancer: laparoscopy or laparotomy? *Gynecol Oncol.* 2000;78(3 Part 1):329–35.
45. Eltabbakh GH, Shamonki MI, Moody JM, Garafano LL. Laparoscopy as the primary modality for the treatment of women with endometrial carcinoma. *Cancer.* 2001;91(2):378–87.
46. Evans KD. A cost utility analysis of sonohysterography compared with hysteroscopic evaluation for dysfunctional uterine bleeding. *J Diagn Med Sonogr.* 2000;16(2):68–72.
47. Fernandez H, Kobelt G, Gervaise A. Economic evaluation of three surgical interventions for menorrhagia. *Hum Reprod.* 2003;18(3):583–7.
48. Guo Y, Longo CJ, Xie R, Wen SW, Walker MC, Smith GN. Cost-effectiveness of transdermal nitroglycerin use for preterm labor. *Value Health.* 2011;14(2):240–6.
49. Horowitz NS, Gibb RK, Menegakis NE, Mutch DG, Rader JS, Herzog TJ. Utility and cost-effectiveness of preoperative autologous blood donation in gynecologic and gynecologic oncology patients. *Obstet Gynecol.* 2002;99(5 Pt 1):771–6.
50. Jack SA, Cooper KG, Seymour J, Graham W, Fitzmaurice A, Perez J. A randomised controlled trial of microwave endometrial ablation without endometrial preparation in the outpatient setting: patient acceptability, treatment outcome and costs. *BJOG Int J Obstet Gynaecol.* 2005;112(8):1109–16.
51. Jakovljevic M, Varjadic M, Jankovic SM. Cost-effectiveness of ritodrine and fenoterol for treatment of preterm labor in a low-middle-income country: a case study. *Value Health.* 2008;11(2):149–53.
52. Kilonzo MM, Sambrook AM, Cook JA, Campbell MK, Cooper KG. A cost-utility analysis of microwave endometrial ablation versus thermal balloon endometrial ablation. *Value Health.* 2010;13(5):528–34.
53. Kovac SR. Decision-directed hysterectomy: a possible approach to improve medical and economic outcomes. *Int J Gynecol Obstet.* 2000;71(2):159–69.
54. Lain SJ, Roberts CL, Bond DM, Smith J, Morris JM. An economic evaluation of planned immediate versus delayed birth for preterm prelabour rupture of membranes: findings from the PPROMT randomised controlled trial. *BJOG: Int J Obstet Gynaecol.* 2017;124(4):623–30.
55. Lalchandani S, Baxter A, Phillips K. Is helium thermal coagulator therapy for the treatment of women with minimal to moderate endometriosis cost-effective: a prospective randomised controlled trial. *Gynecol Surg.* 2005;2(4):255–8.
56. Lenihan JP, Kovanda C, Cammarano C. Comparison of laparoscopic-assisted vaginal hysterectomy with traditional hysterectomy for cost-effectiveness to employers. *Am J Obstet Gynecol.* 2004;190(6):1714–20.
57. Liem SM, van Baaren GJ, Delemarre FM, Evers IM, Kleiverda G, van Loon AJ, et al. Economic analysis of use of pessary to prevent preterm birth in women with multiple pregnancy (ProTWIN trial). *Ultrasound Obstet Gynecol.* 2014;44(3):338–45.
58. Lumsden MA, Twaddle S, Hawthorn R, Traynor I, Gilmore D, Davis J, et al. A randomised comparison and economic evaluation of laparoscopic-assisted hysterectomy and abdominal hysterectomy. *Br J Obstet Gynaecol.* 2000;107(11):1386–91.
59. Marino P, Houvenaeghel G, Narducci F, Boyer-Chammard A, Ferron G, Uzan C, et al. Cost-effectiveness of conventional vs robotic-assisted laparoscopy in gynecologic oncologic indications. *Int J Gynecol Cancer: Off J Int Gynecol Cancer Soc.* 2015;25(6):1102–8.
60. Morrison JC, Chauhan SP, Carroll CS Sr, Bofill JA, Magann EF. Continuous subcutaneous terbutaline administration prolongs pregnancy after recurrent preterm labor. *Am J Obstet Gynecol.* 2003;188(6):1460–5 (**discussion 5–7**).
61. Niinimäki M, Karinen P, Hartikainen AL, Pouta A. Treating miscarriages: a randomised study of cost-effectiveness in medical or surgical choice. *BJOG: Int J Obstet Gynaecol.* 2009;116(7):984–90.

62. Palomba S, Russo T, Falbo A, Manguso F, D'Alessandro P, Mattei A, et al. Laparoscopic uterine nerve ablation versus vaginal uterosacral ligament resection in postmenopausal women with intractable midline chronic pelvic pain: a randomized study. *Eur J Obstet Gynecol Reprod Biol.* 2006;129(1):84–91.
63. Petrou S, Taher SE, Abangma G, Eddama O, Bennett P. Cost-effectiveness analysis of prostaglandin E2 gel for the induction of labour at term. *BJOG Int J Obstet Gynaecol.* 2011;118(6):726–34.
64. Petrou S, Trinder J, Brocklehurst P, Smith L. Economic evaluation of alternative management methods of first-trimester miscarriage based on results from the MIST trial. *BJOG Int J Obstet Gynaecol.* 2006;113:879–89.
65. Prick BW, Duvekot JJ, van der Moer PE, van Gemund N, van der Salm PC, Jansen AJ, et al. Cost-effectiveness of red blood cell transfusion vs. non-intervention in women with acute anaemia after postpartum haemorrhage. *Vox Sanguinis.* 2014;107(4):381–8.
66. Ramsey PS, Harris DY, Ogburn PL Jr, Heise RH, Magtibay PM, Ramin KD. Comparative efficacy and cost of the prostaglandin analogs dinoprostone and misoprostol as labor preinduction agents. *Am J Obstet Gynecol.* 2003;188(2):560–5.
67. Relph S, Bell A, Sivashanmugarajan V, Munro K, Chigwidden K, Lloyd S, et al. Cost effectiveness of enhanced recovery after surgery programme for vaginal hysterectomy: a comparison of pre and post-implementation expenditures. *Int J Health Plan Manag.* 2014;29(4):399–406.
68. Sarlos D, Kots L, Stevanovic N, Schaer G. Robotic hysterectomy versus conventional laparoscopic hysterectomy: outcome and cost analyses of a matched case-control study. *Eur J Obstet Gynecol Reprod Biol.* 2010;150(1):92–6.
69. Sculpher M, Manca A, Abbott J, Fountain J, Mason S, Garry R. Cost effectiveness analysis of laparoscopic hysterectomy compared with standard hysterectomy: results from a randomised trial. *BMJ.* 2004;328:134–7.
70. Sculpher M, Thompson E, Brown J, Garry R. A cost effectiveness analysis of goserelin compared with danazol as endometrial thinning agents. *Br J Obstet Gynaecol.* 2000;107(3):340–6.
71. Simon J, Gray A, Duley L. Cost-effectiveness of prophylactic magnesium sulphate for 9996 women with pre-eclampsia from 33 countries: economic evaluation of the Magpie Trial. *BJOG: Int J Obstet Gynaecol.* 2006;113(2):144–51.
72. Sjöström S, Kopp Kallner H, Simeonova E, Madestam A, Gemzell-Danielsson K. Medical abortion provided by nurse-midwives or physicians in a high resource setting: a cost-effectiveness analysis. *PLoS One.* 2016;11(6):e0158645.
73. ten Eikelder MLG, van Baaren G-J, Rengerink KO, Jozwiak M, de Leeuw JW, Kleiverda G, et al. Comparing induction of labour with oral misoprostol or Foley catheter at term: cost effectiveness analysis of a randomised controlled multi-centre non-inferiority trial. *BJOG: Int J Obstet Gy.* 2017. doi:[10.1111/1471-0528.14706](https://doi.org/10.1111/1471-0528.14706).
74. van Baaren G-J, Broekhuijsen K, van Pampus MG, Ganzevoort W, Sikkema JM, Woiski MD, et al. An economic analysis of immediate delivery and expectant monitoring in women with hypertensive disorders of pregnancy, between 34 and 37 weeks of gestation (HYPITAT-II). *BJOG: Int J Obstet Gynaecol.* 2017;124:453–61.
75. van Baaren GJ, Jozwiak M, Opmeer BC, Oude Rengerink K, Benthem M, Dijksterhuis MG, et al. Cost-effectiveness of induction of labour at term with a Foley catheter compared to vaginal prostaglandin E(2) gel (PROBAAT trial). *BJOG: Int J Obstet Gynaecol.* 2013;120(8):987–95.
76. Vijgen SM, Koopmans CM, Opmeer BC, Groen H, Bijlenga D, Aarnoudse JG, et al. An economic analysis of induction of labour and expectant monitoring in women with gestational hypertension or pre-eclampsia at term (HYPITAT trial). *BJOG: Int J Obstet Gynaecol.* 2010;117(13):1577–85.
77. Walker KF, Dritsaki M, Bugg G, Macpherson M, McCormick C, Grace N, et al. Labour induction near term for women aged 35 or over: an economic evaluation. *BJOG Int J Obstet Gynaecol.* 2017;124(6):929–34.
78. Yoong W, Fadel MG, Walker S, Williams S, Subba B. Retrospective cohort study to assess outcomes, cost-effectiveness, and patient satisfaction in primary vaginal ovarian cystectomy versus the laparoscopic approach. *J Minim Invas Gynecol.* 2016;23(2):252–6.
79. Bijen CB, Vermeulen KM, Mourits MJ, Arts HJ, Ter Brugge HG, van der Sijde R, et al. Cost effectiveness of laparoscopy versus laparotomy in early stage endometrial cancer: a randomised trial. *Gynecol Oncol.* 2011;121(1):76–82.
80. van Dongen JM. Economic evaluations of worksite health promotion programs [Doctoral dissertation]. Amsterdam: Vrije Universiteit Amsterdam; 2014.
81. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Stat Med.* 2000;19(23):3219–36.
82. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess.* 1999;3(2):1–134.
83. Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clin Trials.* 2007;4(2):154–61.
84. MacNeil Vroomen J, Eekhout I, Dijkgraaf MG, van Hout H, de Rooij SE, Heymans MW, et al. Multiple imputation strategies for zero-inflated cost data in economic evaluations: which method works best? *Eur J Health Econ.* 2016;17(8):939–50.
85. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care.* 2005;21(2):240–5.
86. Norwegian Medicines Agency. Guidelines on how to conduct Pharmacoeconomic Analyses. Oslo, Norway: Norwegian Medicines Agency; 2012. p. 77.
87. Collège des Economistes de la Santé. French Guidelines for the Economic Evaluation of Healthcare Technologies. Paris, France: Collège des Economistes de la Santé; 2004. p. 90.
88. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating Medical Costs from Incomplete Follow-Up Data. *Biometrics.* 1997;53(2):419.