Human Reproduction, Vol.27, No.12 pp. 3460–3466, 2012
Advanced Access publication on October 2, 2012 doi:10.1093/humrep/des342

human reproduction

REVIEW *Infertility*

Methodological quality of systematic reviews in subfertility: a comparison of Cochrane and non-Cochrane systematic reviews in assisted reproductive technologies

B. Windsor[1], I. Popovich[1], V. Jordan[2,3], M. Showell[3], B. Shea[4], and C. Farquhar[2,3,*]

[1]Faculty of Medical and Health Sciences, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand [2]Department of Obstetrics and Gynaecology, New Zealand Branch of the Australasian Cochrane Centre, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand [3]Department of Obstetrics and Gynaecology, Cochrane Menstrual Disorders and Subfertility Group, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand [4]Community Information Epidemiological Technologies, University of Ottawa, Ottawa, Ontario, Canada

*Correspondence address. Tel: +64-09-9239493; Fax: +64-09-3035969; E-mail: c.farquhar@auckland.ac.nz

Submitted on June 29, 2012; resubmitted on August 20, 2012; accepted on August 23, 2012

STUDY QUESTION: Are there differences in the methodological quality of Cochrane systematic reviews (CRs) and non-Cochrane systematic reviews (NCRs) of assisted reproductive technologies?

SUMMARY ANSWER: CRs on assisted reproduction are of higher methodological quality than similar reviews published in other journals.

WHAT IS KNOWN ALREADY: The quality of systematic reviews varies.

STUDY DESIGN, SIZE AND DURATION: This was a cross-sectional study of 30 CR and 30 NCR systematic reviews that were randomly selected from the eligible reviews identified from a literature search for the years 2007–2011.

MATERIALS, SETTING AND METHODS: We extracted data on the reporting and methodological characteristics of the included systematic reviews. We assessed the methodological quality of the reviews using the 11-domain Measurement Tool to Assess the Methodological Quality of Systematic Reviews (AMSTAR) tool and subsequently compared CR and NCR systematic reviews.

MAIN RESULTS AND THE ROLE OF CHANCE: The AMSTAR quality assessment found that CRs were superior to NCRs. For 10 of 11 AMSTAR domains, the requirements were met in >50% of CRs, but only 4 of 11 domains showed requirements being met in >50% of NCRs. The strengths of CRs are the *a priori* study design, comprehensive literature search, explicit lists of included and excluded studies and assessments of internal validity. Significant failings in the CRs were found in duplicate study selection and data extraction (67% meeting requirements), assessment for publication bias (53% meeting requirements) and reporting of conflicts of interest (47% meeting requirements). NCRs were more likely to contain methodological weaknesses as the majority of the domains showed <40% of reviews meeting requirements, e.g. *a priori* study design (17%), duplicate study selection and data extraction (17%), assessment of study quality (27%), study quality in the formulation of conclusions (23%) and reporting of conflict of interests (10%).

LIMITATIONS, REASONS FOR CAUTION: The AMSTAR assessment can only judge what is reported by authors. Although two of the five authors are involved in the production of CRs, the risk of bias was reduced by not involving these authors in the assessment of the systematic review quality.

WIDER IMPLICATIONS OF THE FINDINGS: Not all systematic reviews are equal. The reader needs to consider the quality of the systematic review when they consider the results and the conclusions of a systematic review.

STUDY FUNDING/COMPETING INTEREST(S): There are no conflicts with any commercial organization. Funding was provided for the students by the summer studentship programme of the Faculty of Medical and Health Sciences of the University of Auckland.

Key words: assisted reproduction / subfertility / systematic reviews / AMSTAR

Introduction

Systematic reviews and meta-analysis are used increasingly to guide decisions in health care. For the field of subfertility and its associated high cost of treatment and moderate success rates (at best only 20–30% live births per cycle started), the stakes are high (Australian and New Zealand Assisted Reproduction Database, 2008). Therefore, it is important that all steps in an assisted reproduction cycle are based on the best available evidence. Systematic reviews attempt to provide the best available evidence by providing pooled results from studies that answer a specific research question and meet pre-specified eligibility criteria. Systematic reviews have explicit methods that allow the review to be reproducible, and include assessments of the validity of the included studies and a synthesis of the results, usually with meta-analyses. These methods minimize bias, so that more accurate and reliable conclusions can be made (Egger *et al.*, 2001).

Keeping up-to-date with subfertility literature is a considerable task, and clinicians and policymakers should not be expected to examine individual clinical trials to find guidelines or a recommendation for their practice. Systematic reviews and meta-analyses are generally considered the best evidence for synthesizing trials on a specific research question. This superiority is attributed to a formulated review question, predefined eligibility criteria, comprehensive search strategies for trials and assessment of the methodological quality of these trials (Egger *et al.*, 2001). However, if these reviews are not well conducted, then there is the possibility of bias leading to incorrect conclusions (Egger *et al.*, 2001). Systematic reviews need to follow a systematic process so that this bias is minimized.

The Preferred Reporting Items for Systematic reviews and Meta-analyses (PRISMA) statement provides a 27-item checklist and a four-phase flow diagram to assist authors in improving the reporting of quality of their reviews (Moher *et al.*, 2009). This statement is also provided as a resource to Cochrane systematic review (CR) authors. The CR Collaboration supports the reporting items in the PRISMA statement. CR editorial groups use an internal checklist that reflects the content of PRISMA (Green *et al.*, 2011).

The CR Menstrual Disorders and Subfertility Group (MDSG) is one of 53 CR groups and has over 40 CRs published in the CR Library on the different stages of an assisted reproduction cycle. Currently, there are also over 150 systematic reviews of assisted reproduction published in regular peer-reviewed journals. The objective of this study was to compare the quality of the reviews on the CR Library with the non-Cochrane systematic reviews (NCRs).

Materials and Methods

We used the definition of systematic reviews used by the CR Collaboration and in the PRISMA statement: 'A systematic review is a review of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyze and summarize the results of the included studies. Meta-analysis refers to the use of statistical techniques in a systematic review to integrate the results of included studies' (Egger *et al.*, 2001).

Study selection

There were 30 CRs and 30 NCRs included in this study. Using a random number generator, we randomly selected 30 of each type of review from the corresponding list of published subfertility reviews on interventions for assisted reproductive technologies (ART). Each review was numbered (for CRs from 1 to 41 and for NCRs from 1 to 127), and then a random number generator was used to choose a study. Following this, the review was critiqued and included if it met the inclusion criteria. If it did not meet the criteria, then another number was generated and the same process was repeated. Disagreements were resolved by discussion or consultation with a third author (V.J.). The NCRs included in this study were found by searching the PROCITE database of NCRs that is maintained by the Trials Search Coordinator of the MDSG. This database was searched using the following keywords in the titles or lists of keywords: 'IVF' or '*in vitro* fertilization' or '*in vitro* fertilization' or 'ICSI' or 'intracytoplasmic sperm injection' or 'Embryo' or '*in vitro* fertilization' or 'preimplantation genetic screening' or 'assisted reproductive technologies (ARTs)' or 'assisted reproductive technology'.

Only reviews about assisted reproduction were included. Both sets of reviews were searched for the years 2007–2011. There was no language restriction and each review had to include at least one study. Exclusions were reviews looking at diagnostic outcomes, reviews of IUI and reviews of surgery for infertility.

Characteristics of included systematic reviews

Two authors (I.P. and B.W.) independently extracted, onto a spreadsheet, the characteristics and epidemiology of the included studies using the full-text versions. Disagreements between the authors were solved by discussion and consensus or by consulting a third party (V.J.). Authors extracted characteristics that included general characteristics, types of ART intervention, search strategies, risk of bias assessment and statistical analyses.

Methodological quality of included systematic reviews

The Measurement Tool to Assess the Methodological Quality of Systematic Reviews (AMSTAR) was the chosen quality assessment tool for this study. Two authors independently assessed the methodological quality of the included reviews using the AMSTAR tool. This tool was developed (Shea *et al.*, 2007a) and validated (Shea *et al.*, 2007b) in 2007 and assesses 11 aspects of methodological quality with the aim of critically appraising systematic reviews (Table I). Domains were occasionally given a 'not applicable' score, for example, domain 9 if there was only one study in each subgroup so that no pooling was possible.

Statistical analysis

Frequency data are presented as numbers and percentages. Continuous data are presented as medians and inter quartile range (IQR). Risk ratios (RR) and 95% confidence intervals (CI) have been calculated using the fixed effects model and *P*-values are given.

Results

Search results

The initial search returned 168 citations, of which 41 were CRs and 127 were NCRs. Inspection of the titles and abstracts identified 140 potentially eligible reviews that went through a process of random selection to create a data set of 30 CRs and 30 NCRs. After retrieving the full-text versions of each of these reviews, a further nine reviews were excluded and replaced by random selection. A total of 60 systematic reviews were confirmed for assessment (Supplementary data Files 1 and 2). The reasons for exclusion are given in Fig. 1.

Table I Description of the domains of AMSTAR.

Domain	Description
1	Was an 'a priori' design provided?
2	Was there duplicate study selection and data extraction?
3	Was a comprehensive literature search performed?
4	Was the status of publication (i.e. grey literature) used as an inclusion criterion?
5	Was a list of studies (included and excluded) provided?
6	Were the characteristics of the included studies provided?
7	Was the scientific quality of the included studies assessed and documented?
8	Was the scientific quality of the included studies used appropriately in formulating conclusions?
9	Were the methods used to combine the findings of studies appropriate?
10	Was the likelihood of publication bias assessed?
11	Was the conflict of interest stated?

Description of included studies

Table II reports on the journals in which the NCRs were published . Table III lists the descriptive features of the included reviews. The systematic reviews identified covered every stage of an ART cycle. Half of the reviews considered pharmacological interventions (57% of CRs and 47% of NCRs). The stages of ART addressed most frequently concerned the pre-ART phase and down-regulation or ovarian stimulation. Live birth rate was assessed in most CRs (83%) and in two-thirds of the NCRs (67%). CRs were more likely than NCRs to report harms, including ovarian hyperstimulation syndrome (OHSS; 50 versus 20%) and miscarriage rates (73 versus 43%).

The methodological characteristics (not including the AMSTAR domains) of the included reviews are shown in Table IV. Of the included reviews, CRs were more likely to report the inclusion criteria and the research question using the patient, intervention, comparison and outcome (PICO) format than NCRs [RR 1.36 (95% CI 1.09, 1.69) and RR 1.49 (95% CI 1.15, 1.92) respectively]. CRs searched a median of eight databases, which was greater than the NCR median of four. The databases that were used in nearly all the reviews were MEDLINE (100% CRs versus 83% NCRs), EMBASE (100% CRs

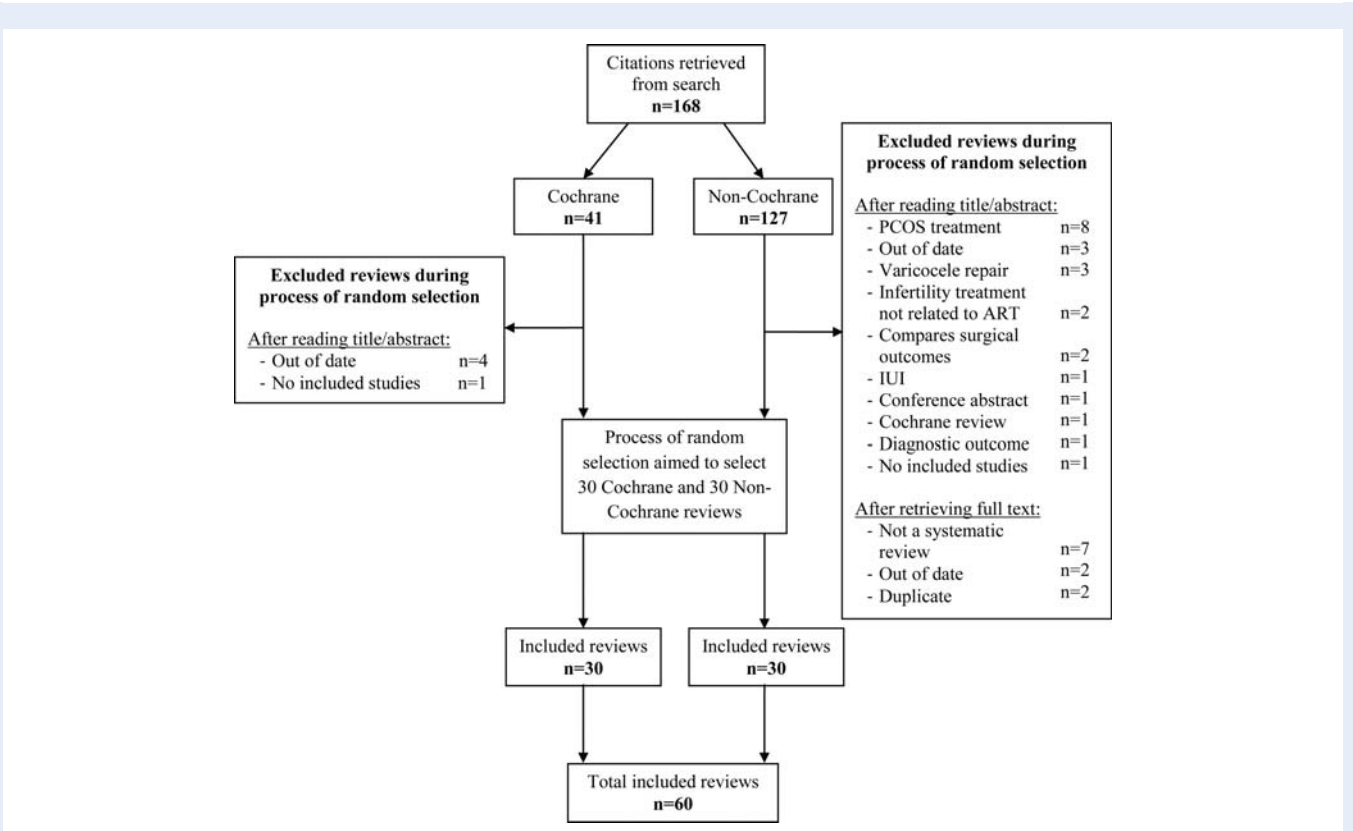

Figure 1 Process of study selection.

versus 83% NCRs) and CENTRAL (97% CRs versus 97% NCRs). The CR MDSG Specialized Register was an additional source of studies for the CRs (100%) but was utilized in only 13% of NCRs. Other sources also searched included reference lists (100% CRs versus 80% NCRs), conference proceedings/abstracts (77% CRs versus 47% NCRs) and hand searching journals (33% CRs versus 0% NCRs). CRs were more likely to include studies identified by hand searching than NCRs [RR 10.0 (95% CI 1.36, 73.3)].

Table II Journals publishing the non-Cochrane reviews.

Journal	Number
Human Reproduction Update	5
Fertility and Sterility	4
Human Reproduction	4
Reproductive BioMedicine Online	4
Reproductive Biology and Endocrinology	3
British Medical Journal	2
Bio Drugs	1
British Journal of Obstetrics and Gynaecology	1
Current Opinion in Obstetrics and Gynecology	1
Expert Review of Obstetrics and Gynecology	1
Gynecologic and Obstetric Investigation	1
Gynecological Endocrinology	1
Human Fertility	1
Journal of Assisted Reproduction and Genetics	1

Heterogeneity was investigated in all CRs and in 90% of NCRs. Of reviews with a minimum of 10 included studies, 53% of CRs used a funnel plot to assess publication bias, while NCRs used funnel plots in 27% and statistical tests in 20% of reviews.

Most CRs reported their funding sources, whereas only one-third of NCRs had a similar statement [RR 2.78 (95% CI 1.57, 4.91)]. No CR received funding from a for-profit industry. However, among the NCRs that reported funding, equal numbers reported non-profit, for-profit and no funding sources. Within the reviews, half of CRs and one-third of NCRs reported the funding of the individual primary inclusion studies. General reports of conflict of interest were always given in CRs but were addressed in less than half of NCRs (47%) [RR 2.10 (95% CI 1.44, 3.07)].

AMSTAR assessment

Figure 2 outlines the percentage of CRs and NCRs that were given a score of yes for each domain. For 10 of 11 AMSTAR domains, the requirements were met in >50% of CRs (domains 1–10), but only 4 of 11 domains showed requirements being met in >50% of NCRs (domains 4, 5, 6 and 9). The strengths of CRs are the *a priori* study design, comprehensive literature search, explicit lists of included and excluded studies and assessments of internal validity. The most significant failings in the CRs were found in domains 2 (duplicate study selection and data extraction in 67% of reviews), 10 (assessment of publication bias in 53% of reviews) and 11 (reporting of conflict of interest: 47%). NCRs were more likely to contain methodological weaknesses as the majority of the domains showed requirements being met in <40% of reviews for example, domains 1: (*a priori*

Table III Description of interventions and outcomes of included reviews.

Category	Characteristics	Cochrane reviews (n = 30)		Non-Cochrane reviews (n = 30)		P-value
Intervention *n* (%)	Pharmacological	17	(56.7)	14	(46.7)	
	Non-pharmacological conventional	12	(40.0)	13	(43.3)	
	Complementary and alternative medicine treatment	1	(3.3)	3	(10.0)	
Stage of ART *n* (%)	Adjuncts and other measures during and pre-ART	4	(13.3)	7	(23.3)	
	Down-regulation/ovarian stimulation	8	(26.7)	7	(23.3)	
	Ovulation triggering	2	(6.7)	0	(0)	
	Sperm/oocyte retrieval	2	(6.7)	0	(0)	
	Lab procedures	1	(3.3)	5	(16.7)	
	Embryo transfer	6	(20.0)	5	(16.7)	
	Luteal phase support	1	(3.3)	3	(10.0)	
	Prevention of OHSS[a]	3	(10.0)	1	(3.3)	
	Special populations	3	(10.0)	2	(6.7)	
Benefits reported *n* (%)	Live birth rate	25	(83.3)	20	(66.7)	0.14
Harms reported[b] *n* (%)	OHSS[a]	15	(50.0)	6	(20.0)	0.01
	Miscarriage	22	(73.3)	13	(43.3)	0.02
	Multiple pregnancy	19	(63.3)	7	(23.3)	0.002
	Cycle cancellation	6	(20.0)	2	(6.7)	0.13
	Fetal abnormality	4	(13.3)	1	(3.3)	0.16
	Ectopic pregnancy	7	(23.3)	2	(6.7)	0.07
	No harms assessed	2	(6.7)	8	(26.7)	0.04

[a]OHSS, ovarian hyperstimulation syndrome.
[b]Does not equal 100% as some reviews reported multiple harms.

Table IV Description of methodology of included reviews.

	Cochrane reviews (n = 30)		Non-Cochrane reviews (n = 30)		Risk ratio (95% CI)	P-value
General characteristics						
Year published median (IQR)	2010	(2008–2011)	2009.5	(2009–2010)	—	
Year of most recent search, if given median (IQR)	2009	(2008–2010)	2008	(2008–2009)	—	
Study design						
Inclusion criteria reported *n* (%)	30	(100)	22	(73)	1.36 (1.09, 1.69)	0.005
PICO research question[a] *n* (%)	30	(100)	20	(67)	1.49 (1.15, 1.92)	0.0001
Search characteristics						
Number of databases searched median (IQR)	8	(6–12.5)	4	(4–6)	—	
e.g. MEDLINE searched *n* (%)	30	(100)	25	(83)	1.20 (1.01, 1.42)	0.04
EMBASE searched *n* (%)	30	(100)	25	(83)	1.20 (1.01, 1.42)	0.04
CENTRAL searched *n* (%)	29	(97)	29	(97)	1.00 (0.91, 1.10)	1.00
Cochrane MDSG Specialised Register *n* (%)	30	(100)	4	(13)	2.26 (1.51, 3.38)	<0.0001
Number of other sources searched median (IQR)	2.5	(2–4)	2	(1–2)	—	—
e.g. Reference lists searched *n* (%)	30	(100)	24	(80)	1.24 (1.03, 1.50)	0.02
Conference abstracts searched *n* (%)	23	(77)	14	(47)	1.64 (1.07, 2.53)	0.02
Journals hand searched *n* (%)	10	(33)	0	(0)	10.00 (1.36, 73.33)	0.0008
Experts/authors contacted *n* (%)	16	(53)	8	(27)	2.00 (1.01, 3.95)	0.04
Years of coverage reported *n* (%)	30	(100)	26	(87)	1.15 (0.99, 1.34)	0.1
Search terms reported as Full Boolean *n* (%)	29	(97)	13	(43)	2.23 (1.47, 3.38)	<0.001
No language restriction, reported[b] *n* (%)	18	(60)	23	(77)	0.78 (0.55, 1.11)	0.26
Results						
Number of studies included/review median (IQR)	10	(4–18)	8	(6–13)	—	
Number of participants/review median (IQR)	1 504	(414–2 508)	1 457	(980–2 544)	—	
Review flow reported *n* (%)	30	(100)	29	(97)	1.03 (0.94, 1.13)	0.47
List of included studies *n* (%)	30	(100)	30	(100)	1.00 (0.94, 1.07)	1.00
List of excluded studies *n* (%)	30	(100)	16	(53)	1.85 (1.32, 2.58)	0.0003
Reasons for exclusion of studies reported *n* (%)	30	(100)	29	(97)	1.00 (0.91, 1.10)	1.00
Heterogeneity investigated (or intent to investigate) *n* (%)	30	(100)	27	(90)	1.11 (0.97, 1.27)	0.13
Pooling of data *n* (%)	27	(90)	28	(93)	0.96 (0.83, 1.12)	0.64
Publication bias assessed (or intent to assess)	16	(53)	11	(37)	1.45 (0.82, 2.59)	0.20
Funnel plot mentioned/assessed *n* (%)	16	(53)	8	(27)	2.00 (1.01, 3.95)	0.05
Statistical tests used *n* (%)	0	(0)	6	(20)	0.08 (0.00, 1.31)	0.02
Other features						
Funding of review reported *n* (%)	25	(83)	9	(30)	2.78 (1.57, 4.91)	0.0004
Non-profit *n* (%)	20	(67)	3	(10)	6.67 (2.21, 20.09)	<0.0001
For-profit *n* (%)	0	(0)	3	(10)	0.14 (0.01, 2.65)	0.24
No funding *n* (%)	5	(17)	3	(10)	1.67 (0.44, 6.36)	0.71
Funding of primary inclusion studies reported	15	(50)	8	(27)	1.88 (0.94, 3.75)	0.08
Conflict of interest of reviewers reported *n* (%)	30	(100)	14	(47)	2.10 (1.44, 3.07)	0.0001
Economics (i.e. costs) considered *n* (%)	7	(23)	8	(27)	0.88 (0.36, 2.11)	0.77

[a]PICO patients, interventions, comparisons and outcomes.
[b]For Cochrane reviews, there were no language restrictions in 18 and it not reported in 12. For non-Cochrane reviews, there 2 reviews restricted to English only, 23 had no language restrictions and it was not reported in 5.

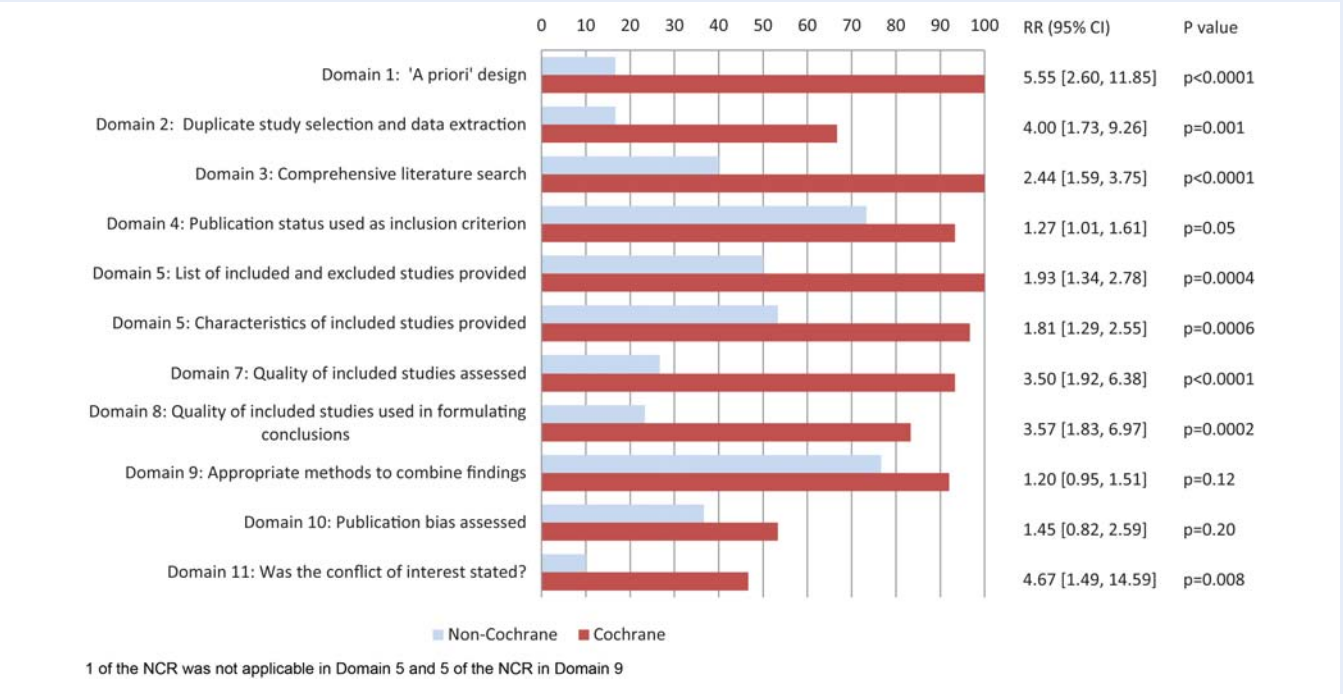

Figure 2 Cochrane versus non-Cochrane: individual AMSTAR scores for each domain given as percentage of reviews receiving a Yes.

study design in 17%), 2 (duplicate study selection and data extraction in 17%), 7 (assessment of study quality in 27%), 8 (including study quality in the formulation of conclusions in 23%) and 11 (reporting of conflict of interest in 10%).

Discussion

This study considered the methodological characteristics and AMSTAR quality assessment of both CRs and NCRs. CRs of ART can be considered to have higher quality than those published in regular journals. In some domains, there were major differences between CRs and NCRs, notably in the pre-stated methods (*a priori* study design) and in assessing study quality and reporting on conflicts of interests.

An *a priori* study design is an important feature in reporting systematic reviews (Green *et al*., 2011). Reviewers need to approach the evidence in a manner that does not allow them to prejudge the outcome of the review. Without a protocol or *a priori* study design, it is possible to change the methods and in particular the analysis and the subgroup analysis. CRs are all preceded by a published protocol that has been peer reviewed. This requires objectives, inclusion criteria and planned analyses and subgroup analyses to be specified in advance (*a priori*) without knowledge of the results of individual trials (Egger *et al*., 2001; Green *et al*., 2011). This approach reduces authors' chances of making decisions about excluding studies that are outliers, or seeking different review outcomes by conducting selective subgroup analyses.

Another area of major difference between CRs and NCRs relates to the assessment of the scientific quality of the included studies and the manner in which the quality of the included studies influenced the validity of conclusions formulated. Assessing study quality is central to systematic review quality and is recommended in the CR Handbook that says the discussion 'should lead to an overall judgment of the internal validity of the results of the review' (Higgins and Green, 2011). All authors of systematic reviews should seek to appraise the risk of bias of the included studies and take those assessments into consideration in the interpretation of the results and the recommendations for treatment.

The funding of the review process and also the funding of the included studies is another domain (11) where differences were present between CRs and NCRs. Domain 11 has two parts - the funding of the included studies and any conflicts of interest that the reviewers may have. While CRs reported their source of funding in 83% of reviews, only 30% of NCRs reported their funding and 10% of the NCRs were funded by for-profit sources. However, only 50% of CRs reported on the funding of the included studies and for NCRs, this was 27%. This is discussed further in the paragraphs below regarding improving the reporting of CRs. All Cochrane authors made statements about conflicts of interest, whereas fewer than half of the NCRs reported on possible conflicts of interest.

The Cochrane Collaboration provides authors with methodological guidance, and therefore the failings in the AMSTAR assessments can be attributed either to the authors not adhering to the guidance, or the guidance not being sufficiently explicit. The domains that need to be addressed by future CR authors are domains 2, 10 and 11. Domain 2 (two independent authors select studies and extract data) should be straightforward for authors to apply but it does require team work. Domain 10 (assessment of publication bias where there are a minimum of 10 studies) again is possible within the CR software

(Revman 5) as a funnel plot can be undertaken. Publication bias (where trials that have positive findings are published more quickly and more often than those with negative findings; Hopewell *et al*., 2009) may result in different conclusions being drawn and recommendations being made by the review authors, and therefore may affect the practice of clinicians. One study found that by including unpublished data with negative findings, the conclusion changed from being favourable to unfavourable for selective serotonin reuptake inhibitors (Whittington *et al*., 2004). Readers need to know that the review authors have considered publication bias, and how it may affect the outcomes of the systematic review.

Finally, domain 11 of AMSTAR assesses the reporting of the funding of all included studies alongside an expression of conflict of interest and funding for the systematic review. Research that is funded by pharmaceutical companies is more likely to publish favourable outcomes supporting the company (Lexchin *et al*., 2003).

This study is limited because the AMSTAR assessment can only make judgments on what is reported by the authors in the publication. Another limitation of this study is that some of the authors include editors and staff of the CR MDSG (C.F. and M.S.) and may have introduced bias. Efforts were made to mitigate this bias by not involving these authors in the assessment of any of the individual systematic reviews. A final limitation was that we were not able to include all the systematic reviews that met our inclusion criteria and, in the case of the NCRs, we have only assessed a representative sample.

This study of a cross-section of systematic reviews on assisted reproduction has reported that CRs have higher methodological quality than similar reviews published in regular journals and can be considered to be the 'gold standard' for clinical decision making. Journals and authors of all journals including the CR Library should consider the PRISMA items in order to improve the quality of their reviews and better inform clinical decision making.

Supplementary data

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

Acknowledgements

We are grateful to Helen Nagels for assisting in editing the manuscript.

Authors' roles

B.W. and I.P. were involved in selecting the studies, extracting the information and entering the data from the studies and in writing the draft manuscript. M.S. conducted the searches of the literature to identify systematic reviews of assisted reproduction and edited the manuscript. V.J. was the final arbiter if there were disagreements between the first two authors and also was involved in the study design and editing. C.F. and B.S. were involved in all aspects of the study design, statistical analysis and preparation of the manuscript.

Funding

Funding was provided for the students by the summer studentship programme of the Faculty of Medical and Health Sciences of the University of Auckland.

Conflict of interest

None of the authors have conflicts of interest to do with commercial organizations.

C.F. and M.S. are both involved in the Cochrane Menstrual Disorders and Subfertility Group. C.F. is the Co-ordinating Editor and M.S. is a Trial Search Coordinator and reviewer. V.J. is employed as the New Zealand Cochrane Fellow of the New Zealand Branch of the Australasian Cochrane Centre. B.S. is the author of the original AMSTAR publication.

References

Australian and New Zealand Assisted Reproduction Database. http://www.acaivf.com.au/ Assisted reproductive technology in Australia and New Zealand. http://www.aihw.gov.au/publications/index.cfm/title/10753 (2008).

Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care*, 2nd edn. London, UK: BMJ Books, 2001.

Green S, Higgins JPT, Alderson P, Clarke M, Mulrow CD, Oxman AD. Chapter 1: Introduction. In Higgins JPT, Green S (eds). *Cochrane Handbook of Systematic Reviews of Interventions*. Chichester, UK: John Wiley & Sons, 2011, 3–9.

Higgins JPT, Green S. Chapter 4: Guide to the contents of a Cochrane protocol and review. In Higgins JPT, Green S (eds). *Cochrane Handbook of Systematic Reviews of Interventions*. Chichester, UK: John Wiley & Sons, 2011, 51–79.

Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009:MR000006.

Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *Br Med J* 2003;**326**:1167–1170.

Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;**62**:1006–1012.

Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007a;**7**:10.

Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, Ramsay T, Bai A, Shukla VK, Grimshaw JM. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One* 2007b;**2**:e1350.

Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet* 2004;**363**:1341–1345.