SHORT REVIEW

Systematic evaluation of the quality of clinical practice guidelines on the use of assisted reproductive techniques

ROSA BERTHA GUTARRA-VILCHEZ1, LETICIA BARAJAS-NAVA2, ALICIA ALEMAN3, IVAN SOLÁ2, IGNASI GICH1, XAVIER BONFILL2 & PABLO ALONSO-COELLO2

1Vitarte Hospital, Health Ministry, Lima, Perú, 2Centro Cochrane Iberoamericano Biomedical Research Institute (IIB-Sant Pau), Hospital de la Santa Creu i Sant Pau, Barcelona, Spain, 3Departamento de Medicina Preventiva y Social, Facultad de Medicina de la Universidad de la República de Uruguay, Montevideo, Uruguay, and 4CIBER de Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain

Abstract

Objective: To conduct a systematic evaluation of clinical practice guidelines (CPGs) on the use of assisted reproductive technologies.

Methods: We searched Medline, the Turning Research into Practice database, and guidelines-specific databases from December 2006 to November 2011. Three reviewers independently assessed each Guideline using the Appraisal of Guidelines for Research Evaluation (AGREE) II instrument. A standardized score was calculated separately for each of the six domains.

Results: Fourteen Guidelines were included. Overall, the quality of these was suboptimal. The scores for each AGREE II domain ranged between 37% and 80%. Three (22%) were deemed “Recommended”; nine (64%), “Recommended with modifications”; and two (14%), “Not recommended”. Agreement among reviewers was very good (Intraclass Correlation Coefficient: 0.915 [95% CI 0.807–0.970]).

Conclusions: The overall quality of the CPGs on Assisted Reproduction Techniques published during the last 5 years is suboptimal. Most Guidelines present significant shortcomings in important domains such as “stakeholder involvement”, “rigor of development”, and “applicability”. Instruments such as the AGREE II and “the Grading of Recommendation Assessment Development and Evaluation” system could prove useful to improve CPGs in this field. Guideline users could benefit from the present results when choosing which guidelines to implement.

Keywords: Assisted reproduction techniques, clinical practice guidelines, in vitro fertilization, assisted embryo transfer, AGREE II instrument, quality assessment

Introduction

Assisted Reproduction Technologies (ARTs) provide potential treatments for infertility, which has a reported prevalence ranging from 3.5% to 16.7% in developed countries and from 6.9% to 9.3% in developing countries (Boivin et al., 2007). ART is an overarching term for all treatments or procedures that include in vitro handling of both human oocytes and sperm or of embryos for the purpose of establishing a pregnancy (Zegers-Hochschild et al., 2009). Worldwide, ART has facilitated the conception of around 246,000 babies in a given year, (ICMART et al., 2009) and there has been an increase in the number of ART cycles performed during the last few years. However, the clinical pregnancy rates of ART are low (28.6–33%) (Nygren et al., 2011; de Mouzon et al., 2012), and vary depending on a variety of factors (McLernon et al., 2010; Pinborg et al., 2011; Maheshwari et al., 2011; Marinakis and Nikolaou 2011).

Clinical Practice Guidelines (CPGs) are defined as “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” (Field & Lohr, 1990). Developing CPGs may be an effective approach to improve the results of ART. “CPGs can bridge the gap between the growing stream of scientific evidence and clinical practice by facilitating the choices of health care professionals and patients for the most appropriate, safe, and cost-effective care” (Nelen et al., 2008). CPGs can minimize risk and improve performance of ART. However, they need to be of high quality, requiring not only knowledge of the clinical question, but also expertise in their development (Fervers et al., 2003).

The Appraisal of Guidelines for Research Evaluation (AGREE) is a validated instrument used to evaluate the quality of CPGs (The AGREE Collaboration, 2003). The original AGREE instrument, developed in 2005,
was updated recently and renamed AGREE II (Burls, 2010; Brouwers et al., 2010; Bowers et al., 2010a). A recent systematic review showed that, despite an improvement over time, the quality of CPGs has remained moderate to low during the last 2 decades (Alonso-Coello et al., 2010).

The quality of ART guidelines has not been systematically appraised using the AGREE II instrument in the last 5 years (Nelen et al., 2008). In order to address this gap in knowledge, we conducted a systematic evaluation of CPGs on ART published in the period 2006–2011.

**Methods**

**Searching for clinical practice guidelines**

We searched MEDLINE (accessed by means of PubMed) using a search strategy with free terms and their corresponding MeSH terms (“Reproductive Techniques”, “Fertilization in Vitro”, “Sperm Injections Intracytoplasmic”, “Emryo Transfer”, “Zygote Intrafallopian Transfer”, “Cryopreservation”, and “Oocyte Donation”) limiting the results with a filter. We also searched the TRIP database, websites of professional societies, and guideline-specific databases (Guideline International Network database, National Guidelines Clearinghouse, Guidelines Finder, CMA InfoBase, and National Health and Medical Research Council) using a simple search strategy, and performed complementary searches in Google.

**Eligibility criteria**

CPGs on ARTs written in English and published between December 2006 and November 2011 were included. ARTs, according to the WHO, include (i) in vitro fertilization and embryo transfer, (ii) gamete intrafallopian transfer, (iii) zygote intrafallopian transfer, (iv) tubal embryo transfer, (v) gamete and embryo cryopreservation, (vi) oocyte and embryo donation, and (viii) gestational surrogacy (Zegers-Hochschild et al., 2009). Systematic reviews, documents explicitly classified as expert consensus, and those that are not available in full text were excluded. Additionally, we excluded GPCs that address laboratory-specific issues of organization or those that do not directly address indications of ART. We considered only the latest version of each CPG.

Two reviewers (R.GV, and L.BN.) reviewed the titles and abstracts of the references retrieved in order to verify eligibility. They also independently examined the full-text versions to discover if they could be included. If needed, disagreements were resolved by discussion with a third reviewer (P.AC.).

**Appraisal of methodological quality of the selected CPGs**

Three reviewers (R.GV, L.BN, and A.A.) appraised the methodological quality of the selected guidelines independently using the AGREE II instrument the purpose of which is to provide a validated framework to assess the quality of guidelines. This instrument contains 23 key items organized in six domains, followed by two overall item score (“global assessment”). Each domain comprises a single quality dimension of the guide. Each item of the AGREE II and the two global assessment items are graded using a 7-point scale (from 1 “strongly disagree” to 7 “strongly agree”). Domains are (1) scope and purpose, (2) stakeholder involvement, (3) rigour of development, (4) clarity and presentation, (5) applicability, and (6) editorial independence (Table I).

When evaluating each item within domains, a score of 1 is assigned when there is no information relevant to the AGREE II item or if the concept is very poor. A score of 7 is given if the quality of reporting is exceptional and if the full criteria and considerations articulated in the User’s Manual have been met. A score between 2 and 6 is assigned when the reporting of the AGREE II item does not meet the full criteria or addresses each consideration. For the overall judgement, a 3-point scale is used to rank the CPG as “recommended” (above 60%), “recommended with modifications” (30–60%), or “not recommended” (below 30%). Domain scores were calculated by adding the scores of all individual items in a domain and by standardizing the total as a percentage of the maximum possible score for that domain (thus ranging from 0 to 100%) (Burls, 2010).

Statistical analyses

We performed a descriptive statistical analysis for each domain. Descriptive values included mean, standard deviation, and minimum and maximum values with 95% confidence intervals (CI). Categorical variables were calculated using the number of cases and the corresponding percentages. Agreement between the three reviewers was determined using the Intraclass Correlation Coefficient (ICC) with a 95% CI. A standardized score was calculated separately for each of the six domains. On the scale proposed by Landis and Koch, the degree of agreement of the ICC is classified as: < 0: Poor, 0.01–0.20: Mild, 0.21–0.40: Right, 0.41–0.60: Moderate, 0.61–0.80: Substantial, and 0.81–1.00: Very good (Kramer & Feinstein, 1981; Landis & Koch, 1977). Data were analysed using SPSS (19.0) for Windows (Statistical Package for the Social Sciences, Chicago, IL, USA).

**Results**

**Search and selection**

From an initial sample of 574 potentially relevant references, 14 CPGs were included (Figure 1). These documents referred to in vitro fertilization (8, 57%),
embryo transfer (4, 29%), embryo cryopreservation (1, 7%), and embryo donation (1, 7%). They were developed in Canada (6, 43%), United States (4, 30%), United Kingdom (2, 14%), Australia (1, 7%), and in several European countries (1, 7%). Nine (64%) were developed by scientific societies, three (22%) by medical associations, and two (14%) by government agencies (Table II).

Appraisal of the methodological quality of CPGs
The overall and per domain agreement among reviewers for the evaluation with the AGREE II instrument was very good (ICC: 0.915, 95% CI 0.807–0.970).

Scope and purpose (Domain 1)
The first domain refers to the overall purpose of the guide, specific health issues, and the target population (items 1–3). The mean score for this domain was 78% (range: 46–98%). The majority of guidelines 12 (86%) had scores above 60% (range: 67–98%). Two guidelines (14%), scored below 60% (range: 46–56%) (Table III).

Stakeholder involvement (Domain 2)
This domain pertains to whether the CPG development groups included individuals from all relevant stakeholders, if the patients’ views and preferences were taken into consideration, and if target users are clearly defined (items 4–6). The mean score for this domain was 45% (range: 17–94%). Most documents (78%) scored below 60% (range: 17–54%). Patients or patient representatives were involved in the development of three of the guidelines (21%) (Lee et al., 2006; NICE 2009; Teede et al., 2011) (Table III).

Rigour of development (Domain 3)
This domain is crucial as it concerns the process used to gather and synthesize the evidence with which to develop the CPG, the methods used to formulate the recommendations, the effects of the recommendations, the explicit link between the recommendations and the supporting evidence, and the methods followed for reviewing the CPG externally and for updating it (items 7–14). The mean score in this domain was 53% (range: 8–99%). Eight guidelines (57%) scored above 60% (range: 63–99%). Six (42%) scored below 60% (range: 8–58%) (Table III).

Clarity of presentation (Domain 4)
This domain assesses whether recommendations were specific and unambiguous, if they covered the corresponding options for disease management according to its scope, and if key recommendations are easily identifiable (items 15–17). With a mean score of 80% (range: 48–98), it was the domain that showed the least variability. Twelve guidelines (86%) scored above 60% (range: 65–98%). Only two (14%) guidelines scored below 60% (range: 48–59%) (Table III).

Applicability (Domain 5)
This domain assesses whether a CPG describes facilitators and barriers to its application, provides advice on, or tools for, how the recommendations can be put into
practice, if it presents monitoring or auditing criteria, and the cost of implementing the corresponding recommendations (items 18–21). Scores for this domain were the lowest, with a mean score of 37% (range: 7–89%). Two guidelines (15%) scored above 60% (range: 86–89%) (Table III).

**Editorial independence (Domain 6)**

This domain refers to the sources of external funding and the possible conflicts of interest among members of the development group (items 22–23). The mean score was 63%, although with great variability (3–100%). Five CPGs (35%) scored below 60% (3–53%), whereas nine (64%) scored above 60% (range: 64–100%) (Table III).

**Overall assessment**

Our results showed that the highest-quality guidelines were those developed by the European Association of Urology, Polycystic Ovary Syndrome Association of Australia, and National Institute for Health and Care Excellence (NICE). Three CPGs (21%) were deemed “recommended” (NCCWCH/NICE, 2009; Dohle et al., 2010; Teede et al., 2011), nine (64%), “recommended with modification” (JOINT SOGC-CFAS, 2008; Lee et al., 2006; Jarvi et al., 2010; Cutting et al., 2008; American Urological Association (AUA), 2010; Min et al., 2010; Vause et al., 2010; Leyland et al., 2010; Liu & Case, 2011), and 2 (14%), “not recommended” (Practice Committee of the American Society for Reproductive Medicine (ASRM), 2008, 2009) (Table III).

**Discussion**

Our review shows that the methodological quality of ART CPGs is suboptimal and most are either not recommended or recommended with modifications. Several of the AGREE domains, including “stakeholder involvement”, “rigour of development”, and “applicability” had poor scores.

The score in “Rigour of development” (53%) was worryingly low. This domain is considered crucial from a methodological point of view and includes amongst other things, how recommendations are developed and the criteria used to rate the quality of the evidence and the strength of recommendations. The World Health Organization, like many other organizations around the world (including the Cochrane Collaboration, NICE, Scottish Intercollegiate Guidelines Network), has...
Table II. Details of selected guidelines on assisted reproduction techniques.

<table>
<thead>
<tr>
<th>Title</th>
<th>Country</th>
<th>Organization</th>
<th>Author</th>
<th>Publication year</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Society of Clinical Oncology recommendations on fertility preservation in cancer patients.</td>
<td>USA</td>
<td>American Society of Clinical Oncology.</td>
<td>Lee et al.</td>
<td>2006</td>
</tr>
<tr>
<td>Guidelines for the number of embryos to transfer following in vitro fertilization.</td>
<td>Canada</td>
<td>Society of Obstetricians and Gynaecologists of Canada and the Board of the Canadian Fertility and Andrology Society.</td>
<td>JOINT SOGC-CFAS</td>
<td>2008</td>
</tr>
<tr>
<td>Guidelines for gamete and embryo donation: a Practice Committee report.</td>
<td>USA</td>
<td>American Society for Reproductive Medicine; Practice Committee of Society for Assisted Reproductive Technology.</td>
<td>ASRM</td>
<td>2008</td>
</tr>
<tr>
<td>Guidelines on number of embryos transferred. Principio del formulario</td>
<td>USA</td>
<td>American Society for Reproductive Medicine and the Practice Committee of the Society for Assisted Reproductive Technology.</td>
<td>ASRM</td>
<td>2009</td>
</tr>
<tr>
<td>The Management of obstructive azoospermia: AUA Best Practice Statement.</td>
<td>USA</td>
<td>American Urological Association Education and Research, Inc.</td>
<td>AUA</td>
<td>2010</td>
</tr>
<tr>
<td>Elective single embryo transfer following in vitro fertilization.</td>
<td>Canada</td>
<td>Society of Obstetricians and Gynaecologists of Canada and the Board of the Canadian Fertility and Andrology Society.</td>
<td>Min et al.</td>
<td>2010</td>
</tr>
<tr>
<td>Ovulation induction in polycystic ovary syndrome.</td>
<td>Canada</td>
<td>Society of Obstetricians and Gynaecologists of Canada and the Board of the Canadian Fertility and Andrology Society.</td>
<td>Vause et al.</td>
<td>2010</td>
</tr>
<tr>
<td>Advanced reproductive age and fertility.</td>
<td>European</td>
<td>European Association of Urology.</td>
<td>Dohle et al.</td>
<td>2010</td>
</tr>
</tbody>
</table>

JOINT SOGC-CFAS, Joint Society of Obstetricians and Gynaecologists of Canada – Canadian Fertility and Andrology Society; ASRM, American Society for Reproductive Medicine; NICE, National Institute for Clinical Excellence; AUA, American Urological Association.
Table III. Mean standardized score per domain and overall assessment results for each Clinical Practice Guideline.

<table>
<thead>
<tr>
<th>Guidelines Authors</th>
<th>Scope and purpose</th>
<th>Stakeholder involvement</th>
<th>Rigour of development</th>
<th>Clarity and presentation</th>
<th>Applicability</th>
<th>Editorial independence</th>
<th>Overall recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al.</td>
<td>89</td>
<td>89</td>
<td>70</td>
<td>48</td>
<td>32</td>
<td>92</td>
<td>Recommended with modifications</td>
</tr>
<tr>
<td>Cutting et al.</td>
<td>87</td>
<td>39</td>
<td>47</td>
<td>81</td>
<td>42</td>
<td>08</td>
<td>Recommended with modifications</td>
</tr>
<tr>
<td>JOIN SOGC-CFAS</td>
<td>83</td>
<td>30</td>
<td>66</td>
<td>89</td>
<td>31</td>
<td>33</td>
<td>Recommended with modifications</td>
</tr>
<tr>
<td>ASRM</td>
<td>67</td>
<td>17</td>
<td>10</td>
<td>72</td>
<td>07</td>
<td>03</td>
<td>Not recommended</td>
</tr>
<tr>
<td>NICE</td>
<td>89</td>
<td>94</td>
<td>90</td>
<td>91</td>
<td>89</td>
<td>83</td>
<td>Recommended</td>
</tr>
<tr>
<td>ASRM</td>
<td>72</td>
<td>24</td>
<td>08</td>
<td>67</td>
<td>10</td>
<td>93</td>
<td>Not recommended</td>
</tr>
<tr>
<td>AUA</td>
<td>56</td>
<td>28</td>
<td>32</td>
<td>65</td>
<td>32</td>
<td>81</td>
<td>Recommended with modifications</td>
</tr>
<tr>
<td>Min et al.</td>
<td>72</td>
<td>35</td>
<td>67</td>
<td>87</td>
<td>38</td>
<td>64</td>
<td>Recommended with modifications</td>
</tr>
<tr>
<td>Vase et al.</td>
<td>91</td>
<td>31</td>
<td>65</td>
<td>87</td>
<td>38</td>
<td>64</td>
<td>Recommended with modifications</td>
</tr>
<tr>
<td>Leyland et al.</td>
<td>69</td>
<td>31</td>
<td>58</td>
<td>94</td>
<td>33</td>
<td>50</td>
<td>Recommended with modifications</td>
</tr>
<tr>
<td>Jarvi et al.</td>
<td>46</td>
<td>30</td>
<td>17</td>
<td>59</td>
<td>19</td>
<td>83</td>
<td>Recommended with modifications</td>
</tr>
<tr>
<td>Dohle et al.</td>
<td>91</td>
<td>54</td>
<td>63</td>
<td>96</td>
<td>50</td>
<td>100</td>
<td>Recommended</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>85</td>
<td>46</td>
<td>63</td>
<td>91</td>
<td>17</td>
<td>69</td>
<td>Recommended with modifications</td>
</tr>
<tr>
<td>Teed et al.</td>
<td>98</td>
<td>94</td>
<td>99</td>
<td>98</td>
<td>86</td>
<td>100</td>
<td>Recommended</td>
</tr>
</tbody>
</table>

recognized the need to use more rigorous processes to ensure that health care recommendations are based on the best available research evidence (Oxman et al., 2006; Fretheim et al., 2006). The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) can help guideline authors improve their performance in this domain, since it provides a system for rating quality of evidence and strength of recommendations that is explicit, comprehensive, and structured (Guyatt et al., 2008; Schünemann, 2009; Brozek et al., 2011; Guyatt et al., 2011).

“Stakeholder Involvement” also had a low score (45%), probably due to patients or patient representatives not being involved in the development of guidelines. Guideline developers should include important stakeholders such as consumers, relevant health professionals who work within the relevant area, and managers or policy makers. They should also have access to the advice of experts or individuals with the necessary technical skills in the field of the CPG. In addition, draft recommendations should be reviewed by consumers, who should be asked explicitly to consider the values that were used (Fretheim et al., 2006; Schunemann et al., 2006; Cavazos et al., 2008).

Finally, the low score in the domain “applicability” (37%) might indicate that CPG developing groups lack training on strategies to implement CPGs and to determine the resources needed for that end. Previous research has identified physician barriers as the main reason for non-adherence to CPGs. These barriers include lack of familiarity with guidelines, perceived limited validity of guidelines, limited applicability of guidelines among specific patients, clinical inertia, influence of prior anecdotal experiences, medical heuristics, and guideline factors such as unclear or ambiguous guideline recommendations (Cavazos et al., 2008; Lugtenberg et al., 2009). Another study showed that “multiple barriers impede physicians’ adherence to subfertility guidelines, mainly physicians’ lack of self-efficacy and low outcome expectancy” (Haagen et al., 2005). The identification of these factors, which influence the applicability of recommendations, could help developers, and ultimately the implementation and use of recommendations. The ART procedures are not considered by all health insurances but some have suggested that insurance mandates might influence ART utilization. (Sunderam et al., 2012; Butts et al. 2013). National Health systems should be aware of the more trustworthy CPGs if they decide on funding ART.

On the other hand, “Scope and purpose”, “Clarity of presentation”, and “Editorial independence” had reasonably good scores. The CPGs assessed received high scores on how they defined their “scope and purpose”, which indicates that the questions in these CPGs were generally well defined. The score for “Clarity of presentation” domain was above 60% and showed the least variability. The “Editorial independence” domain also had good scores, though with high variability, implying that authors of CPGs tend to provide information about potential conflicts of interest.

These results are broadly consistent with previous evaluations in the field of assisted reproduction (Appleyard et al., 2006; Haagen et al., 2006; Nelen et al., 2008). In a previous review, both the “Scope and purpose” and “Clarity of presentation” domains had the highest scores. Similarly, the scores for “Rigour of development”, “Stakeholder involvement”, and “Applicability” were far from optimal score (Alonso-Coello et al., 2010). In contrast, the highest score in our review was given to “Editorial independence”.

Our evaluation has strengths and limitations. Its strongest point is that three independent evaluators achieved a high degree of agreement when they assessed the CPGs selected. In addition, we independently evaluated the inclusion criteria and implemented a systematic appraisal of the methodological quality of ART guidelines with the recently developed and internationally validated AGREE II instrument. Regarding limitations, we focussed on CPGs that were published in the period 2006–2011 and contained treatment recommendations.
about the use of ART. We therefore may have excluded other high-quality guidelines about other issues around ART (such as ethical, laboratory, and organizational issues, or preventing infertility) or CPGs that had been published earlier. However, the vast majority of CPGs on this and other fields are updated periodically. One intrinsic limitation of the AGREE instrument is that it does not differentiate between “not performed” and “Not noted”. To the extent that guideline groups did some of the things that they do not report we would be underestimating the true quality of the included guidelines. However, we think that this is unlikely. Finally, expert consensus documents were not included as they lack the basics of evidence-based recommendations (e.g. an explicit search strategy and the classification of the quality of the evidence). These documents are likely to be of lower quality than the ones included in our review. To that end, results may overestimate the quality of CPGs in the field of ART, strengthening our conclusions.

In summary, the overall quality of CPGs on ART published in English during the period 2006–2011 is suboptimal. The low score in the “Rigour of development” domain is particularly remarkable, since this domain is considered an important surrogate of methodological quality. It appears that a great deal remains to be done to reach excellence in CPGs in general, and specifically those on ART. Initiatives like GRADE and AGREE II could prove beneficial in improving CPGs’ quality in future. Guideline users could benefit from the present results when choosing which guideline to implement. Similarly, professional organizations that develop CPGs, might take into account these results to improve the domains “stakeholder involvement” and “applicability”.

Acknowledgements
R.GV. is a PhD candidate at the Paediatrics, Obstetrics, and Gynaecology, Preventive Medicine and Public Health Department, Universitat Autònoma de Barcelona, Spain. She obtained a fellowship from the Ford Foundation International Fellowships Program.

We would like to acknowledge Hector Pardo for his help in the edition of the final version of the manuscript.

Declaration of interest: The authors report no declarations of interest. The authors alone are responsible for the content and writing of the paper.

P.A.C. is funded by a Miguel Servet contract by the Instituto de Salud Carlos III (CP09/00137).

References


Quality of clinical practice guidelines on reproductive techniques 7


The biostatistics of concordance. Clinical Pharmacology and Therapeutics, 29, 111–123.


Practice Committee of the American Society for Reproductive Medicine (ASRM); Practice Committee of the Society for Assisted Reproductive Technology. (2008). Guidelines for gamete and embryo donation: A Practice Committee report. Fertility and Sterility, 90, S30–44.


Appendix 1: Full electronic search strategy for MEDLINE


© 2014 The British Fertility Society