Reviewer's report

Title: Difficulty and appropriateness of decision-making by General Practitioners: a systematic review of scenario studies

Version: 2  Date: 4 September 2014

Reviewer: Susan Hrisos

Reviewer's report:

General comments
This is a well written manuscript, though it is somewhat repetitive in places. It addresses an interesting topic, the review methodology appears to be robust and the review process is for the most part very well conducted. The single coder approach to some aspects of the review undermines this to an extent and lessens confidence in the authors’ findings and interpretation. The review set out to answer three research questions: the extent to which decision difficulty and decision appropriateness, and the relationship between them, has been studied; the determinants of decision difficulty and appropriateness; and to investigate the relationship between difficulty and appropriateness of GPs’ decisions.

Minor essential revisions
Page 3 – 4: Opening argument is somewhat repetitive but generally well written
Page 4 lines 86-89 – the proposition that decision making could be improved through the use of ‘fast and frugal’ heuristics and ‘ignoring part of the information’ seems to be presuming the results before the investigation – perhaps discussion of this and other possible approaches to improving clinical decision making should be kept for the discussion.
Page 5 line 105 – no restrictions on the format and delivery method. Does the manuscript comment on these features and whether or not they make a difference to the review findings?
Page 5 line 106 - Studies using just one scenario were excluded as ‘they couldn’t be compared’. How comparable are the individual scenarios within and across each of the included studies. Is it not valuable to compare what different GPs found easy or difficult across the same scenario?
Line 107 ‘qualitative’ studies were excluded but interview studies were included? This requires clarification.
Line 146 – Decisions made must have been assessed against ‘some standard’ – were these recognised standards and were they national, local etc?
Line 152-153 – Here the authors acknowledge the heterogeneity between studies yet decisions from each were still pooled. It is also unclear about how ‘the number of appropriate decisions was calculated’. Was the number of decisions analysed explicitly mentioned in each paper – if not how were they identified and who identified them?
Line 154 – first mention of decision type – how was this categorised? Describe categorisation first and then mention chi square tests – makes more sense.

Line 157 – Table 1. The categories of decision types and decision quality were agreed amongst all authors but it is a significant limitation that the coding of decisions within these categories was undertaken by only one author, especially given the vast number of individual decisions. Was any validity checking done?

Line 171 – is it not that the studies used questionnaires rather than they were questionnaires? What is a straightforward questionnaire? And how did they differ from those nested within other studies (other than that they were part of other studies)?

Line 175- what is a straightforward interview?

Line 177 how was decision difficulty and appropriateness measured in the interview studies?

Can the authors comment in more detail about the range of scenarios used within studies – how were they constructed and what was their quality? What formats and delivery methods were used?

Line 188 – repetition (key features of studies in appendix 4). Forty-three studies used guidelines – what id the other 19 use? The key features of these 66 studies should be summarised in an embedded table similar format as Table 3.

Major essential revisions

The analysis of decision appropriateness was restricted to data extracted from only 38/152 included studies and represents just over 50% of studies that assessed decision appropriateness. Could the remaining 28 studies not contribute anything to the findings (why no qualitative synthesis of these studies?)

The analysis of decision difficulty is restricted to four studies, though only one lent itself to any attempt to quantify factors that influenced decision difficulty. I have some major reservations about this latter analysis, not least that it was undertaken by a sole reviewer thus increasing the potential for bias. What were the ‘pieces of information’ that were counted and were they specifically selected for inclusion by the original authors? E.g. were the pieces of information combinations of the ‘multiple factors’ mentioned at line 216? Were they all relevant or were some elements noise? Was the information vague or precise and was the distribution of such factors equivalent across scenarios? Did one additional factor really tip the balance to change perceptions from easy to difficult or was it the nature of the information pieces included? I also wonder if some of the difficulty that GPs are attributing to difficulty is a reflection of the lack of the simulation rather than difficulty with the decision per se. For example, they may have several unanswered questions that they would ordinarily ask in the real life situation.

Minor essential revisions

Generally the review findings appear to be confirming what is already known or hypothesised – that increased scenario complexity influences GPs perceptions of
difficulty and that this is associated with poorer or inappropriate decision making. However, the review was unable to expand on this as planned in terms of identifying what other factors contribute to decision difficulty. Perhaps the novel contribution of this review needs to be clarified more prominently for the reader?

Discussion

There is some repetition in this section, introduction of new data (line 269 – the range of appropriate decisions across included studies). This needs clarification - for example does the 6% figure represent the proportion of appropriate decisions made by GPs? Can the authors comment on why there was such a variation across the studies? Is it possible for example that scenario evaluation of decision making becomes less valuable as case complexity increases? Scenarios cannot capture important interactional aspects of the clinical context.

The authors question the extent to which this variation across studies reflects real practice (line 271) but then go on at line 326 to say that the review findings have broad generalisability.

References 24, 26 & 27 that are used to support the authors argument across lines 275 to 284 and some others) appear dated – are there no more recent studies available? How do the authors’ estimates of rates of inappropriate care delivery compare with others’?

Line 280 – only 46% advice giving behaviours were appropriate. The authors suggest that this may indicate that ‘GPs decisions are not solely influenced by the ‘relevant’ evidence base’ and allude to time constraints as potential factors. This surely is not a new observation. There is later some brief discussion about the use of guidelines to reflect real world decision making (line 291) suggesting that there are numerous ways in which health care quality can be conceptualised. This point might be worth expansion.

Line 302 – I think it is a strong claim to say that there was ‘a trend for more participants to perceive scenarios as difficult …’ based on the analysis of the content of five scenarios from a single study. Can we really conclude on the basis of this extremely limited analysis that one additional piece of information increased the complexity of the scenarios sufficiently to shift perceptions from easy to difficult?

Line 324 – here and throughout the discussion, perhaps all the references to ‘further research’ could be summarised under an ‘implications for research’ heading?

Line 324 Strengths and limitations – the study has more limitations than strengths- a summary at the end of this section outlining the implications of these limitations for the interpretation and generalisation of the findings would be beneficial.

Line 349 – potential for double counting. The number of studies that this happened for may be small but what about inflation of the number of appropriate decisions contributing to an analysis?

Line 356 - if scenarios are not valid reflections of real practice how can it be more ‘ethical’ to base improvement strategies on inaccurate assessments of clinical
practice?

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests