Author's response to reviews

Title: Delivering stepped care for depression: an analysis of implementation in routine practice

Authors:

David A Richards (d.a.richards@exeter.ac.uk)
Peter Bower (peter.bower@manchester.ac.uk)
Christina Pagel (c.pagel@ucl.ac.uk)
Alice Weaver (a.e.weaver@exeter.ac.uk)
Martin Utley (m.utley@ucl.ac.uk)
John Cape (John.Cape@candi.nhs.uk)
Steve Pilling (s.pilling@ucl.ac.uk)
Karina Lovell (karina lovell@manchester.ac.uk)
Simon Gilbody (sg519@york.ac.uk)
Judy Leibowitz (judy.leibowitz@camdenpct.nhs.uk)
Lilian Owens (lilianowens@hotmail.com)
Roger Paxton (rspaxton@googlemail.com)
Sue Hennessy (sue.hennessy01@ntlworld.com)
Angela Simpson (aes4@york.ac.uk)
Steve Gallivan (s.gallivan@ucl.ac.uk)
David Tomson (d.p.c.tomson@ncl.ac.uk)

Version: 5 Date: 19 August 2011

Author's response to reviews: see over
Cover Letter 19th August 2011 Revised Manuscript

Dear Professor Kent,

Richards, DA et al: Delivering stepped care for depression: an analysis of implementation in routine practice

Thank you for reviewing the above paper and requesting revisions. We have undertaken these as requested and have detailed them below. Our apologies that the revised manuscript does not come with tracked changes. Because our revision was an almost complete re-write it became impossible to work on using tracked changes and I have not been able to retrieve an up to date tracked changes version. We have, therefore, provided very detailed descriptions of our changes below. I do hope this will allow you and the reviewers to ascertain whether we have improved the manuscript as requested.

Yours sincerely,

David A. Richards, University of Exeter, UK.

Reviewer: Elizabeth Yano

Reviewer’s report:

MAJOR COMPULSORY REVISIONS

1. The utility of defining evidence based medicine and health services research for a journal whose audience is already very familiar with these concepts is limited, and does not substantively contribute to the argument being made for studying implementation of stepped care in the context of NICE guidelines. This section should be significantly shortened to get to the point of depression guidelines sooner and with greater focus and intent. The material in Box 1, in contrast, is essential to the argument being made, and should probably be in the narrative rather than segmented off in a Box (though the Figure within the Box is useful).

We agree with the reviewer in principle, although we hope that the paper will be read by those in the general area of psychotherapy research, as well as those with a specialist interest in implementation. We have retained some of these arguments, but have significantly reduced them in length.

2. The authors suggest that the traditional delivery of care for depression has "performed adequately" and that the main issue is one of "demand for care." This reviewer is not aware of any such evidence that has been population-based in nature, especially in view of the substantial barriers to access to effective depression care management, the stigma associated with seeking/obtaining care in specialty mental health settings, and the high volumes of patients with depression who are seen in primary care rather than mental health settings. The authors need to take better account of the range of depression care modalities and approaches that stem from guidelines.

We do not think there is a fundamental disagreement here, as the issues that the reviewer identifies (barriers to care, stigma) relate to demand and access, and we were trying to highlight the importance of these issues. However, we have removed the sentence that has caused confusion, as it was not critical to our argument, and the referee highlighted that this section need to be reduced.
3. The abstract needs to be rewritten substantively. On the surface of it, it suggests that the paper represents a modest and limited contribution to the literature, presents conclusions that are not described in the results section, and makes broad and generic recommendations. The abstract would benefit from a stronger anchoring in the details of stepped care and the extent to which the NICE guidelines are so specific that the resulting variation observed is problematic vs. reflective of appropriate local adaptations.

This has been completely rewritten as requested to sharpen up the text and make it reflective of the revised manuscript.

4. The paper would benefit from better integration of examples throughout. (e.g., page 8, "steps relating to treatments of different intensity" would benefit from provision of at least one example; page 8, "minimal interventions" should be defined and/or an example given). Clarifying the "tension" between "stepped vs. stratified" and defining how they differ would be important. No references for the "tension" or issues the authors are trying to frame are provided, so the foundation for some of these statements is unclear. Actually, starting the "tension" sentence off as a new paragraph may be helpful and/or defining stepped care carefully and then getting into the importance of the contrast with stratified care may be useful. If the relative importance of the two mechanisms was unclear according to NICE guidelines, why would variations be a surprise in terms of their implementation? And if the evidence per the authors for a decision between the two is limited, isn't this as it should be?

We have used the term 'low intensity' consistently in the revision, and defined it more clearly in the text. We have used a new paragraph to described the tension between 'stepped' and 'stratified' care as suggested, and provided more detail to highlight the differences between these models of care.

We agree that we have given the impression that variation was unexpected, and have modified the text to make it clear that the existence of variation is not such a surprise, but that it highlights the need to use evidence from studies such as these to explore and potentially reduce variation in later iterations of the guidelines.

5. Background, page 9, "the extent to which" -- needs to be refined and clarified as to what the "critical issue" is. Similarly, the authors have not made an adequate case for what the "HTA/HSR effectiveness/delivery dilemma" is. Currently, the questions that the paper aims to answer do not seem particularly well linked to the introductory statements leading up to them. The questions themselves are however reasonably important ones that may provide useful information for certain audiences. For example, the introduction does not make a case for why knowing more about how patients access and flow through stepped care is important (this reviewer thinks it likely is important, but the logical flow in the introduction doesn't get us there).

We have simplified this section, highlighting that stepped care (as defined in the NICE guidelines) underpins significant investment in services in the United Kingdom, and that assessment of the implementation of stepped care is thus critical

6. Methods, page 10, if the authors "assisted sites in designing their stepped care models...", then wouldn't some of the variation be due to their procedures? There seems to be an undercurrent in how the paper is currently framed that "blames" the guidelines for being too specific and not specific enough at the same time, which distracts from what the paper actually may have to offer.
As noted above, we agree that the original paper gave an unfortunate impression of blame to the guidelines and the sites, and we have rewritten the paper to avoid this impression. We have also briefly described the consensus process and made it clear that service design was the sole responsibility of the sites and not influenced by the research team.

7. Methods, page 10, the types of data collected are listed, but they are not linked to any sort of conceptual or theoretical framework that serves as a foundation for determining whether they are the right variables and/or where there may be gaps. Analysis of ‘patient pathways’ using “start and end points” is not clear -- maybe an example would help. How the authors determined step directions should be explained -- again an example may help. Did the authors have any hypotheses about the structure and patient flow patterns going into this work?

We have identified, described and referenced the ‘patient pathway’ as the conceptual framework for our data collection. The methods section has been rewritten to clarify the type of data being collected and how it is related to this conceptual framework.

8. Results, page 14, it may not be uncommon for there to be data acquisition problems in operations projects, however, as written, there is no information provided to interpret “large amounts of data were missing for many variables.” Are missing data chiefly among those variables needed to describe the various samples within and across sites? Or was missing data a theme for all measures, including flow patterns? If so, the utility of the available information is admittedly difficult to interpret, limiting the potential takeaway from what is presented as findings. For example, in the absence of better patient-level data or practice-level patient information, how should readers interpret the variations found? Maybe all of these variations are appropriate, maybe not, but either way, as written, they appear to be a list of different sites’ patterns, and what readers are supposed to take away is not sufficiently clear.

We have clarified that it was in fact routine demographic data which was missing in large amounts, not patient pathway/flow data. We have added the rates of missing data.

9. Results, page 14, “multiple routes with multiple inputs and outcomes” -- again one or more examples would anchor this statement in the narrative (in addition to what is provided in the diagrams).

We have provided a brief description of the range of different pathways.

10. Results, page 15, ratios of intensity -- this section is interesting but requires some anchoring in an example or other information to help readers interpret what a 22:1 ratio of intensity means at a practice and/or provider and/or patient level.

We have re-written this section to provide clarity and provide additional information to help readers understand the statistics.

11. Discussion, page 16, paragraph 2, seems like this should be in the Results section (seems like new information). Actually, much of the Discussion section as written is easier to understand, follow and interpret than much of what is in the Results section. In fact, integration and anchoring of findings in the literature does not start in the Discussion section until page 18. Better integration and
framing of the findings in what else is known about stepped care models and about implementation in general would be useful in the Discussion section.

We have removed the data/results section from the discussion and incorporated it into the results section and have restricted the discussion to framing the findings in what else is known about stepped care models and about implementation in general.

12. Discussion, page 20, the authors suggest that the "most significant implication" of their work is "the failure of guidelines to be implemented in a manner consistent with the authors' intentions." The dilemma here is that they set up the argument up front that the guidelines are non-specific with respect to details of stepped care implementation, so consistency with them would by definition create variability as local providers insert their own interpretation and preferences into how they implement the guidelines.

As noted above, we accept that we have been somewhat inconsistent here, highlighting the lack of specificity in the NICE guidelines in terms of the implementation of stepped care, and then seeming to criticise services for their variability. We have modified the tone of these comments to make it clear that variation is to be expected, and further assessment of these sites (and others) may provide useful evidence to feed into later iterations of the guidelines. Our core statement here is that it is important that evidence from implementation studies such as this need to feed into the guideline development process to make a more explicit and effective link between health technology assessment, and research in service delivery and organisation.

MINOR ESSENTIAL REVISIONS

1. Background, end of first paragraph, "consider these..." does not fit the rest of the sentence. Greater clarity in the meaning of "wider service context" is needed.

These phrases have now been removed, following the reviewer's request to shorten the introduction

2. Background, page 7, bottom paragraph, "as" missing between "seen" and "a reasoned."

We have made the requested change

3. Methods, page 10, "four volunteer NHS providers" -- as Implem Sci is an international journal, it would be helpful to clarify early whether this means four individual clinicians or four individual practices. It is not clear in the 2nd paragraph which it is (e.g., "assisted sites"). While this becomes clear under Results, it should be made clear under Methods as well.

We have made the requested change

4. Methods, page 11, "anonymised data were submitted..." This section needs to be clarified with respect to the steps between identifying data from existing information systems vs. those collected by clinicians. To whom were data submitted? And how? And how were data anonymised? Not sure if a table that clarifies these distinctions and procedures would be helpful, but as written, it is unclear what data were obtained from which sources by whom.

We have made it clear that anonymisation was done by local staff before being passed to the research team. We have also clarified which data were collected by clinicians in addition to the routine systems.
5. Results, page 13, again examples would help readers better understand study findings (e.g., "low intensity treatment steps").

We have made the requested change

6. Results, page 13, please define the "voluntary sector" for your non-UK colleagues. What is a "senior mental health worker"? Please describe the "self-directed computerised treatment programme" more.

We have made the requested changes

Reviewer: Annemieke van Straten

Reviewer's report:

Providing services along Stepped care principles is rapidly gaining popularity worldwide. Even though, as the authors argue, evidence for service delivery (how should it be offered?) is scarce or absent. This paper demonstrates very well what happens when implementing stepped care in the absence of such guidelines. It is one of the first papers demonstrating large practice variations in stepped care.

The main weakness of this paper is the lack of outcome data. Even though this is already acknowledged by the authors in their discussion section they might elaborate a bit more on this crucial point. e.g. I would like to know: is this data not collected at all in the services? or were these data just not available for research purposes? In case it is not collected at all it could be argued that none of these services actually implemented stepped care?

We agree with the reviewer that one of the fundamental principles of stepped care – the assessment of treatment effect to aid clinical decision making appears deficient in these services. However, the use of formal psychometric measures is not the only way to assess treatment progress. Clinical judgement may be applied by mental health workers using clinical interview assessments. Clearly, for research purposes this is less easy to quantify without routine outcome measures but their absence does not necessarily invalidate the stepped care process given our patient flow data. Whether the use of such measures within a formal clinical decision making algorithm would have reduced the variation in patient pathways could be a potential subject for a stepped care research programme. We have added this point to the discussion in the limitations section.

Another point which should be discussed more clearly is the way the stepped care was initially designed by the different services. In their methods section the authors state that they were involved in these designs. I therefore wonder what their role was (and to what extent they were responsible for the ‘failure of guidelines to be implemented in a manner consistent with the authors’ intentions.’ Page 20)? E.g. why did the different services came up with different models? Did the authors aim to create similar designs across services? Maybe some of the services were more ‘stepped care’ minded from the start than others? What did the services do to implement the design they had chosen? Was the staff trained and / or ….? Were the authors involved in this? Were there important differences between the services in the time, energy and money spent on actual implementation?
This comment relates to those made by reviewer 1, where we may have been unclear in our presentation, and suggested that we were being critical of sites, even though there was a lack of clarity in the guidelines, and were part of the implementation process. We have made changes to the manuscript to avoid this impression, and to instead highlight that variation at this stage is to be expected, but that it is important that evidence from implementation studies such as this feed into the guideline development process to make a better link between research in health technology assessment and in service delivery and organisation. We have also clarified our role in the process of service design.

Table 2 shows us that there are 3 major steps within the stepped care design: guided self-help, short term face-to-face interventions, and more complex / specialized face-to-face interventions (next to assessment and crisis interventions). However, part of the paper is about ‘low’ and ‘high’ intensity treatments. I wonder if the authors regard the short face-to-face interventions as ‘low’ or ‘high’ intensity? This should be defined somewhere. 
In my opinion Figure 3 and Figure 4 are not the most interesting. They do make you realize that the patient flow is diverse but nothing else. I think it would be more interesting to put more essential data in a bar chart. Like: the percentage of patients starting with a low intensity treatment, or percentage of patients stepping up.

We agree that our terms ‘low’ and ‘high’ intensity do not map on well to the three major steps, and that this can be confusing. For the purposes of the paper, we define ‘low intensity’ in terms of guided self help (step 2 in table 2), as this is most often what is meant by the term, with short term face-to-face interventions, and more complex / specialized face-to-face interventions both examples of ‘high intensity’. We have made this more explicit in the table and the text.

Although we accept that Figures 2 and 3 are complex, we do feel that they do provide a good visual (and perhaps rather startling) illustration of the different systems implemented in the sites in the way a table, and would prefer to keep them in the revision. We have added the bar chart as an additional figure as requested.
Minor points

The title should reflect that this paper is on delivery of mental health care (and not health care in general)

We have made the requested change

For non-UK readers the authors need to explain the difference between ‘specialized mental health trust’ and ‘primary mental health teams’ (page 12).

We have made the requested change

Also: it is not clear whether the studied services were the only services available for GPs to refer to? Or are there other services and is only a selection of patients referred to the studied services?

This is a limitation and we have referred to it in the limitations section

The authors should explain more clearly what is meant by ‘para-professional mental health (or graduate) workers’ (page 12 and 13).

We have clarified the use of these terms.

The authors argue that ‘large’ amount of data were missing. I would like to have some estimate of this proportion.
We have responded to this in our rewrite of point 8, reviewer 1. We have clarified that it was in fact routine demographic data which was missing in large amounts, not patient pathway/flow data. We have added the rates of missing data.

The patient flow shows us that about 75% of the referred patients are assessed. I guess that this means that the other patients are referred but don’t show up for their appointment in the service? Or are they screened by the service and ‘sent back’ to their GP? Or …???

These patients were not assessed at all. We have moved this information from the discussion section to the results.

The authors show on page 15 that unscheduled discontinuation rates are 43% for low-intensity treatments and 10% for high intensity treatments (might this also be due to selective missing data?) but on page 19 they show that only a third of the high-intensity treatments were completed. This is a bit confusing. And this is an important point since it is often thought that drop-out rates for low intensity treatments are much higher than for high intensity treatments. I wonder if this is demonstrated in this study or not?

We agree that this is confusing and a product of patients with unknown final dispositions. We have removed this from the results and discussion since the confound in this data is unhelpful and the cross sectional nature of the data leaves many patients still receiving treatment at the census date. We have made this point in the discussion.

Table 1: the abbreviation IMD should be explained.

We have made the requested amendment