Author's response to reviews

Title: Depression and Decision-making Capacity for Treatment and Research: A Systematic Review

Authors:

Thomas Hindmarch (thomas.hindmarch@kcl.ac.uk)
Matthew Hotopf (matthew.hotopf@kcl.ac.uk)
Gareth S Owen (gareth.1.owen@kcl.ac.uk)

Version: 2 Date: 27 April 2013

Author's response to reviews: see over
Dear Mr Adrian Aldcroft,

Re. Depression and Decision-making Capacity for Treatment and Research: A Systematic Review. Hindmarch et al. MS: 1287795028436042

Many thanks for your response to this submission and for the broadly positive reviewer comments that we found very helpful. We have made every effort to respond to the criticisms in a point-by-point manner and have substantially revised the paper in consequence. We think the paper is now improved and suitable for publication in BMC Medical Ethics. Please find our responses below.

Yours sincerely

Thomas Hindmarch and Gareth Owen

Reviewer's report

Title: Depression and Decision-making Capacity: A Systematic Review

Version: 1 Date: 5 December 2012 Reviewer: SY Kim

Reviewer's report:

The authors reviewed published clinical ethical analyses and empirical studies of DMC and depression. This is a useful review because there has not been a systematic review of this topic, and it does come up frequently in variety of discussions about DMC. Also, it is quite difficult to do a systematic search of such studies because there is not a uniform categorization scheme in the bibliographic databases. So this would be a valuable contribution.

The authors reviewed two types of literature: ‘clinical ethical accounts’ and more systematic empirical papers that attempted to measure the impact of depression on capacity.

I enjoyed reading this summary, as it nicely puts key papers into one discussion. I tend to agree with the authors that many of the ‘objections’ to appreciation standard
as ‘too cognitive’ are probably not accurate and that the appreciation standard can reasonably be interpreted to capture many of the arguments to the effect that the current standards are too cognitive.

This paper is likely to acceptable for publication but I do think that some corrections need to be made, clarifications added, and some points addressed. I have tried to explain each point below.

Firstly, thank you for taking time to review the paper so thoroughly. Your input has been invaluable and as you will see, we’ve amended the paper based on your critiques and those from the other reviewer. Comments in relation to your individual critiques are below

1. In regard to the first type of papers, the list did seem fairly complete but I did wonder how reproducible the search for those papers would be. How does one decide something has “substantial ethical analysis”? For instance, there were a number of articles in 1990s or so on ‘rational suicide’ focusing on suboptimal choices made by patients due to influence of depression (similar in spirit to the Leeman, Halpern etc papers cited in this review), except those papers in general did not discuss the phenomena under the rubric of ‘capacity’ and thus are not included in this review. I couldn’t help wishing for a more “operational” definition of procedures used to arrive at their list of articles of “clinical ethical analyses.” (Also, it is not clear to me that given the type of content they are seeking, why it should be limited to articles—I’d imagine there are chapters in ethics books that provide rich and useful discussions.)

You quite correctly raise a number of issues here. In response we’ve made some alterations:

a) We have sought to better define what we meant by ‘substantial ethical analysis’. In truth, when reading the selected papers, the difference between an article in which the author(s) had considered the problem of conceptual/ethical connections between DMC and depression and those which had not was pretty clear. We have attempted to make this more transparent in the methodology section using examples. None of the papers excluded by the first author where then included by the last author on review so we are pretty confident we are not excluding a large, significant literature.

b) The mentioned papers on ‘rational suicide’, are definitely relevant for the medical ethics of depression but we have excluded them because they are not directly related to the modern (functional) concept of DMC, tending to relate more to older outcome or status models of capacity/competence.

c) We agree that a formal review of all bioethics books is a limitation, and this is now acknowledged in the discussion. We have conducted an informal review of some of the leading bioethical publications (e.g Beauchamp and Childress’s books, Buchanan and Brock’s book on deciding for others, Carl Elliot’s publications and various anthologies of bioethics) and these haven’t been high yield making us think that a formal review would not add a great deal. We have found books typically offer no, or very cursory, coverage or refer to papers in the review. We add a short discussion of this.

2. In regard to the list of empirical studies:

a. First, do the Stacey et al and the Simon et al studies conform to the authors’ definitions in the methods section? It appears that no “DMC assessment tools” are used in those studies, nor are there “performance measures” of DMC in those
studies. They do have something to do with the depressed patients' opinions about medical decision-making but that is distinct from assessing their abilities. I would recommend against including them in the review. (In fact, a better case can be made to include studies by Ganzini and Lee on severely depressed elderly who change their minds on treatment preferences after treatment of depression, as being more relevant than these two articles... although I don’t think the G&L study fits into this review either since it is not about DMC specifically... my point is that comparatively, they would be more informative than the Simon and Stacey studies).

We agree and have removed these papers from the review.

b. Second, the authors should double check their review details. For example, the Appelbaum MacCAT-CR study of 1999 involved moderately depressed outpatients in a psychotherapy trial, not severely depressed inpatients.

Another example: the 1995 Grisso-Appelbaum paper did not use MacCAT-T as the MS text suggests (although, inconsistently but correctly, in table 2, it is said that precursors to MacCAT-T are used). Also, I’m not sure if you can make the assertion as is done in Table 2 that appreciation is most impaired. In a sample of 92, it's a difference between, for example, 8% with impaired reasoning and 12% with impairment in appreciation, which is not significantly different.

These two articles I am very familiar with, so I happened to see these discrepancies. I would suggest double-checking the summaries of the other articles also.

Thank you for spotting these mistakes. We have re-read all the papers and attempted to correct all the mistakes/add relevant details.

c. Third, perhaps the studies should be summarized with a bit more detail—quantitative details when available, and providing key methodological details when relevant. This seems necessary given the diversity of methods used by these studies, and given that interpretation will depend on patient characteristics, methods of measurement used, and methods of categorizing incapacity. Where ratings were done to measure severity of depression, the summary statistics should be given. The actual frequencies of impaired capacity should be given as well (rather than mere qualitative summaries when in fact summary numbers are available). Also, it is important for the readers to know how the papers categorized their subjects—e.g., the 1995 Grisso study used a statistical cutoff to determine impaired capacity status, and this has an impact on how to interpret the results. Another important factor is actually whether studies like Bean et al measure appreciation at all, given the format of their questions in the instrument. This is important because, as the authors note, methods varied and given the importance that Grisso/Appelbaum have placed on first person versus third person formulations of interview questions for distinguishing between understanding something intellectually versus applying it to oneself, it may be worth noting that not all instruments follow that convention.

We agree. Table 2 has been expanded, giving the reader more detail on study participants, depression type and severity, capacity decision, measures used, study objectives and quantitative findings.

3. Some comments on interpretation of studies reviewed and conclusions drawn.

a. Judgments of incapacity are meant to justify seeking decisional authority in a surrogate. One could have impaired abilities for decision-making but still be capable. This distinction is not made anywhere in the paper, leaving the
impression that any time depression affects decision-making, this implies incapacity.

*We agree with you here and so have added this into the background section of the paper and brought this distinction into the substance of table 2.*

b. The authors assert that measurements of appreciation are deficient bc it is said to be based on detecting incapacity in psychotic patients and not in depressed patients (i.e., MS states, “The law has not articulated its perspective on appreciation...” But law is pretty silent on details of all of these standards.) I’m not so sure about this as a matter of legal history as the authors argue. Perhaps it is, but I think more evidence needs to be provided in support.

Also, an alternative explanation could be that perhaps appreciation standard is by design not meant to be overly sensitive, in the interests of patient autonomy? In the clinic and in the hospital, we often deal with poor choices of patients, but we deal with it clinically rather than resorting to questioning their decisional authority every time.

*These are subtle points and we know the reviewer has thought carefully and deeply about them.*

We agree that the law is fairly silent on many points of detail with regard to the DMC standards and have added this to the text. The legal review by Berg et al. (1996) examining the US cases that were important in the legal construction of the appreciation standard, refers only to the clinical terms “delusion” and “schizophrenia.” We have tried to clarify this point more. Also, according to our understanding of the US legal history, it was the right to refuse anti-psychotic medication that was a driver for the development of the DMC construct in the US (and therefore for the MacCAT-T). We have added this point. So the DMC construct would appear to have grown on medico-legal soil fairly distant to depression. The difficulties the architects of the MacCAT-T had in operationalizing appreciation for depression are recalled in Grisso and Appelbaum’s publication cited. We think this comes out in the MacCAT-T scoring rules for Appreciation where there is the note: “failures to acknowledge the potential benefit of treatment may obtain a 0 rating not only if they are based on delusional belief systems, but also if they are strongly influence by extremes in affective symptoms: e.g., mania, severe depression.” (underline added). (Grisso and Appelbaum 1998 OUP, p187). So depression is recognized in relation to zero scores on appreciation even without delusion but there is the problem of specifying and interpreting what constitutes “strongly influenced”. This will be a problem where hopelessness rather than delusional premises or a serious distortion of reality are the overt clinical features as is often the case in severe depression.

*We agree there are societal issues about where the threshold for DMC should be set and we agree that clinicians often deal with poor choices of patients clinically through support rather than by resorting to formal DMC considerations and that too much formality may have unintended consequences. We have tried to add these points.*

c. From an empirical point of view, I think it is difficult to assert based on current evidence that depression per se has a strong impact on capacity (in the sense of justifying depriving the patient of decisional authority). I worry that a very general
statement like “DMC may be impaired in depressive illness” gives the wrong impression that DMC is routinely impaired in depression, when in fact evidence is to the contrary (indeed, most studies of even severely depressed patients show relatively weak impact).

Of course, the common concern that depression can make people’s decisions worse in some measurable sense is probably true. Severe depression, especially, likely has an effect. However, it is a normative question whether society should deprive a person’s right to make decisions based on the fear that they are not at their best. I get the feeling that the authors may be in favor of lowering the threshold for incapacity based on the appreciation standard (or rather, parts of the paper read in that direction). But it may be that the studies tend to show low rates of incapacity because they operationalize appreciation in a way that reflects the high regard put on autonomy.

We agree the wording here risks creating a misleading impression. We have removed the general sentence “DMC may be impaired in depressive illness” from the abstract and added more specific statements.

Concerning the normative question. We take the point about the high social regard put on a negative concept of autonomy (individual freedom from state interference) but we think the reviewer would agree society can have valid concern about the positive concept of autonomy in people suffering from depression (individual freedom from psychopathology). In other words, ensuring the health interests of such individuals are safeguarded in systems of medical care. We think there is an issue not with lowering the threshold for incapacity but making the test less silent on, and less at risk of arbitrary application in, depression – a form of psychopathology to which the test does apply.

Overall, I think this paper would make a useful contribution to the literature. I hope the above suggestions will help improve it.

We thank you again and hope we have managed to improve this paper.

Level of interest: An article of importance in its field

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'
Reviewer's report

Title: Depression and Decision-making Capacity: A Systematic Review

Version: 1 Date: 20 February 2013

Reviewer: Bettina von Helversen

Reviewer's report:

Before I start I would like to point out that my expertise is not in the field of clinical psychology/psychiatry but cognitive psychology and decision-making. Thus I’m probably not the typical audience BMC Medical Ethics is directed at and some of my comments may only matter to the degree that the article is directed at a wider audience.

That stated, I think the topic of the manuscript is very relevant. I’m still surprised how little empirical research there is investigating how depression affects the ability to make sound decisions and how heterogeneous the approaches to the topic are. Thus this review comes at the right time. However, I have severe reservation about the description of the method and the interpretation and presentation of the evidence. Some may be caused by my problems understanding the article, possibly because I’m not a clinical psychologist and thus not familiar with the tools used to assess competency to make treatment decisions.

Firstly, thank you for reviewing this paper and the attention paid to methods/interpretation. We found the comments helpful and have done our best to improve the paper based on the points you make. Responses to your individual concerns are below:

I explain my concerns in more detail below:

Major Compulsory revisions

1) I had problems following the methods used to select the articles. The inclusion criteria should be explained in more detail. For instance, it should be stated explicitly what are relevant observational data, what kinds of assessment tools were accepted (and which were/would be rejected). Similarly, it should be explained what requirements need to be met for a substantial ethical analysis.

We agree. The description of methods has been revised. Observational data had to relate to legal/ethical standards of DMC and this is stated. This is a relatively new area for empirical research and so we included all assessments tools that related to the legal standards. The MacCAT, however, dominates the empirical literature and is the most validated measure. 2 observational studies have been excluded in response to reviewer 1’s comments that the measures (supported decision making) do not relate directly to DMC. We have added details of measurement to table 2. We have further described how we extracted papers of ‘significant ethical analysis’ by using examples (see also response to reviewer 1).

2. The authors state that one of the main goals of the articles is to review the performance of measures of DMC in depression. But then it is left unclear what these measures are and how the performance of these measures can be evaluated. If this is the goal, this needs to be clarified substantially and the empirical studies analyzed accordingly. I have to admit I’m skeptically this goal can be achieved given the little
research on the topic. A goal which maybe more appropriate given the scarcity of research (and I would still find interesting and relevant) would be to analyze in how far the empirical evidence supports the conclusion that depressed patients are incompetent to make treatment decisions. And if this is the case, what is the proportion of depressed patients that are incompetent, what are the main underlying reasons for the incompetency, and what symptoms of depression may be most predictive of being incompetent. In the discussion these findings could be connected with the more theoretical/ethical analyses of the single case studies.

A couple of issues are raised here and we will attempt to clarify them:

- The ‘measures of DMC’ are referred to throughout the paper and within Table 2. Most studies utilize the MacCAT (either the MacCAT-T for treatment decisions or the MacCAT-CR for research decisions). We’ve added a figure to add more detail on the MacCAT to help the non-specialist reader. There are different ways of measuring DMC (the measure of the abilities and the measure of the judgement is a key distinction) and we’ve added more detail to the text and to table 2 to clarify this.
- In terms of the goals you have suggested, we have had to summarize the data as we find it. This does include data on impairments in the DMC abilities, the proportion of depressed patients judged to lack DMC in different settings and some of the clinical associations (e.g. insight). In essence, the review aims to draw together the available data and interpret it. We agree that there is relatively little research on the topic and we think the reviewer is justified to be surprised by this.

3) I do not see the value of a systematic review of single case studies and ethical analyses. In a similar vein, it was also unclear to me how the difference between an opinion piece and a case study was defined. It seems to me that a case study focusing on one or two depressed patients whose ability to make treatment decisions was impaired is essentially an opinion piece and does not reflect on the probability that any depressed patient will actually suffer from the possible impairment. A systematic review of single case studies suggests that the number of articles is related to the probability of a depressed patient being incompetent to make a treatment decision. Instead of a systematic review I would suggest that the arguments why depressed patients may be incompetent to make treatment decision that are outlined in the single case studies and ethical analyses are summarized in the introduction of the manuscript, but organized by theory/argument and not by article/author.

We think there may be an issue about kinds of legitimate method here. Case study methodology does not use measurements of DMC and probability samples. It uses cases to try to analyse conceptual and meaningful relations between depression and DMC and come up with constructions that capture those relations. That is important in a review on depression and DMC because a) the abilities and judgements in question are normative (moral) so a solely metric approach to the review can be criticized for being inappropriate or one sided (indeed, some people working in this area would solely value a case study approach) and b) in research areas where measures are relatively blunt or approximate one needs to include studies that generate hypotheses about what needs to be measured. We think that clinical ethical analysis using case material does have value and that it can be identified and reviewed. The suggestion of referring to this literature in the introduction we don’t think will be a sufficient or transparent way of identifying all of the material and laying it out systematically for appraisal.

We hope the additions to the methodology section, using examples, clarifies how
clinical ethical studies were selected. The authors did agreement checks on papers excluded and there were no “grey” papers. This has been added. We found that the clinical ethical studies were not as difficult to isolate as may appear.

4) I could not follow the interpretation of the results and the conclusions drawn. Looking at the evidence reported in the empirical articles, it seems to indicate that the majority of depressed patient is not found incompetent, but there is a sizable minority of 15-30% if patients that is. The single case studies/theoretical analyses seem to suggest a more evasive problem than found in the empirical studies. The authors argue that this difference is due to a measurement problem. I don’t follow this conclusion. In my opinion it is equally possible that the clinical analyses are not reflecting the majority of depressed patients or that appreciation is not necessarily the main problem of depressed patients. In particular since the physicians seem to consider even fewer patients to be incompetent than the MacCat assessment tool. This should be discussed in more detail.

We have revised table 2 and parts of the results and discussion sections and abstract. By putting more data in table 2, we hope that the results of each study and our consequent interpretation is clearer. We agree that the majority of depressed patients are found to be competent by measures like structured clinical judgement using the MacCAT-T though in some severely depressed samples it is a considerable percentage. In interpreting the data it is important to note that the MacCAT-T scores do not, by themselves, give a competency judgment.

The clinical ethical analyses are not reporting prevalence of problems with appreciation - simply instances where depression and appreciation are argued to be conceptually connected. We agree that it is possible the clinical ethical analyses have insufficient case material and overemphasize appreciation over other abilities and this mentioned in our conclusion. However, the empirical data is picking up problems with appreciation, especially in more severely depressed groups. There are difficulties with measurement of appreciation in depressed samples using the MacCAT that are discussed. Overall we think the best interpretation of the literature is that appreciation is an important variable in DMC assessment in depression though this does not mean that lack of DMC in depression is common.

Nevertheless, the claim that the assessment tools are not well equipped to measure appreciation in depressed patients could very well be true. However, if the authors want to argue this, they need to provide the reader with much more information about what defines appreciation, how the MacCat tool measures it and what would be other ways of measuring it. For instance, it was not clear to me what kind of behavior would constitute an impairment according to the MacCat tool in any of the four categories. It also did not become clear to me why the MacCat is not a good measure of appreciation.

In the introduction we define appreciation using the standard formulation and it is expanded upon in the first section of the clinical ethical analysis in results. For further clarity we have added information in a figure on the MacCAT now added and references to seminal publications in the area are given. Table 2 now gives more information on MacCAT appreciation scores. In the discussion we discuss appreciation and insight (concepts that are often considered similar) and the problems with using insight as an indicator of appreciation in depression. We don’t argue that MacCAT appreciation does not recognize the impact of depression on appreciation – it does. We argue that it leaves a lot vague. Finding alternative ways to specify and measure appreciation in depression we conclude is an important
research problem (though not an easy one). We don’t think this can be an aim for this literature review.

5) In the section on Empirical Studies, Patient perceptions it was not clear to me whether the patients were still depressed or questioned after they had recovered from a depressive episode. Although patients’ self report are important, I’m also not sure that they are the best way to measure decision making capacities as the authors seem to suggest. If the self reports are collected while people are depressed, their negativity could be a symptom of the depression. However, if they are collected retrospectively it is difficult to assess in how far they reflect the actual state of mind at the decision time or a retrospective construction biased by the knowledge of the depressive episode.

We agree that it is better to remove the paper from the review and have done so. See also response to reviewer 1.

Minor essential revisions

1) In the empirical research it would be nice if in addition to the main finding it was reported if and how many depressed patients were considered as being incompetent to make decisions.

Where reported data allow this is now added.

2) The authors frequently refer to decision making capacities of depressed patients in general. However, in fact they only consider a very specific decision making competency, that is making treatment decisions. This focus should be made clearer (in particular in the title and in the abstract) because now the title is confusing and it’s not clear how the results would generalize to other types of decision tasks.

We consider treatment and research decisions and have made changes in the title and abstract to clarify this. Thank you for this point.

Minor issues not for publication

1) Empirical studies fourth paragraphs (p 10) –the abbreviations ECT should be explained

This has been added.

2) Empirical studies: summary Bean et al.: I did not understand what “discriminated judgment of DMC most” means? Were depressed judged as incompetent according to the tool or to the physician? Did judgments of physicians and the tool differ most on this item?

Physicians made the DMC judgments and CIS scores (measured independently) were studied as associations. One item is reported as discriminating capacity/incapacity most strongly using the Wilks Lambda test. This detail is now added to table 2.

3) Why is Owen et al., 2008 not included in the review?

We agree it should be and have added this to the review. We have also added Cohen et al. 2004 – a paper that on a further check of those selected for detailed
abstract analysis we realized needed to be included.

4) In the discussion in the second to last paragraph, the authors write “we have to consider...”. It was unclear to me if the authors refer to themselves or to doctors in general. Please specify.

We have revised this phrase and hope it is now clearer.

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