Author’s response to reviews

Title: What are the barriers and facilitators for third sector organisations (non-profits) to evaluate their services? A systematic review

Authors:

Anders Bach-Mortensen (anders.bach-mortensen@wolfson.ox.ac.uk)

Paul Montgomery (p.x.montgomery@bham.ac.uk)

Version: 1 Date: 28 Dec 2017

Author’s response to reviews:

Dear Dr Abbott,

Thank you for the decision letter, and for the opportunity to revise our manuscript for publication in Systematic Reviews. We are most grateful for the very helpful comments provided by the reviewers, and feel they have greatly strengthened the paper. Below we specify how we addressed each of the reviewers’ comments.

We look forward to hearing from you soon and look forward to the next stage.

Best wishes for the holidays.

Anders Bach-Mortensen

Paul Montgomery
Reply to reviewers

Reviewer #1:

I thought this review was conducted in a straightforward and sensible manner. I have some only some minor comments:

We thank the reviewer for this comment and for the helpful suggestions below.

83-85 "many third sector organisations (TSOs) fail to evaluate their activities following the principles of evidence-based practice (EBP, which are broadly defined as "

the integration of clinical expertise, patient values, and the best evidence into the decision making process for patient care."" - these principles seem to relate to knowledge translation rather than to be principles for evaluation as such. The same tensions is present in the following lines (84-86): adherence to best practice in delivery of services vs being able to evidence impact. I can see that evaluation would be considered part of this approach (you have to know whether something is working as intended in order to learn from it and to improve things) but you might want to spell this out more explicitly.

In response to this comment and minor comment #3 by Reviewer 2, we have removed the conventional definition of EBP, as we can see it may be misleading in this case. Instead we have framed the sentence to focus on the importance of evaluation. It now reads:

“However, despite the increasing emphasis on impact and performance evaluation by government and commissioners, many third sector organisations (TSOs) fail to evaluate their activities following the principles of evidence-based practice (EBP), which mandate the adherence to rigorous, reproducible and systematic methods [9]. While it is generally acknowledged that monitoring and evaluation according to best practice is important in the delivery of services, there has been little systematic effort to investigate why the third sector is struggling to evidence its impact, despite becoming a growing provider of social and health services [10–13].“ (p4, para 1)
118 "why TSOs struggle to demonstrate their impact" - though could also be that they (or some of them) choose not to do this or do not see the value of it; as it is your wording is slightly normative.

We have changed the wording, so that it becomes clear that we only refer to those TSOs that undertake evaluation. It now reads:

“However, few efforts have addressed why TSOs that engage in evaluation struggle to demonstrate their impact to stakeholders despite the importance of understanding effectiveness (or possible harms) to vulnerable service-users.” (p 5, lines 125-127)

157-8 I thought the wording of the objective was unclear and could be improved. "What barriers and facilitators do third sector organisations experience when evaluating their services as identified by third sector practitioners?" seems to imply that the services are identified by practitioners. I think a better wording would be:

"What barriers and facilitators do third sector practitioners identify in relation to evaluating the services their organisations provide?"

We thank the reviewer for this comment and have amended as suggested.

Reviewer #2:

Thank you for submitting the paper. I am pleased to have been invited to review it, and have found it interesting and enjoyable to read. Please accept my apologies for the delay in providing a report.

Overall assessment:

This is a very good paper. It is mostly well-written, straightforward and clearly presented and structured. I learnt a lot through reviewing it, and, if it is published, the authors will have done a
very useful job of assessing literature in this otherwise fragmented area. It provides a helpful and up to date ground-clearing service, and, sets the scene well for further research. Subject to the comments below, I believe it is publishable, but have four main suggestions designed to improve the paper – though I should stress: I do not think this requires a major re-write. In addition, I have made a more substantial number of (usually very) minor suggestions as a result of a close read of the paper, which are listed in bullet point form at the end of this review.

We thank the reviewer for these comments and for the very helpful suggestions below.

The four main things I would suggest the authors attend to in a revised submission are:

1. Note and utilise the distinction between 'capacity' and 'capability'

The phrase 'lack of expertise and internal capacity' (p.13, lines 327-329) involves, strictly speaking, two separate concepts. The former is about 'capability' (a reference to skill or expertise - 'know-how', or how well something can be done) whereas the latter is 'capacity' (a volume conception of how much or how little something can be done). 'Lack of financial resources' is a capacity issue rather than a capability issue (as is 'lack of time' - p.14, line 341). At the risk of self-referencing, the distinction between them is discussed in Macmillan, R. and Ellis-Paine, A., with Kara, H., Dayson, C., Sanderson, E. and Wells, P. (2014) Building Capabilities in the Voluntary Sector: What the evidence tells us (TSRC Research Report 125, Birmingham, Third Sector Research Centre)(see, e.g. pp.6-7), informed by the Big Lottery Fund's thinking in this area (see: https://www.biglotteryfund.org.uk/research/making-the-most-of-funding/building-capacity). The distinction is well represented in the two main thematic groupings subsequently used in the paper, i.e. 'Factors related to lack of resources' refers to lack of capacity, whilst 'Factors related to technical capacity and evaluation skills' refers to 'capability.' The suggestion is to make use of this distinction in discussing the analysis and results of the review.

We thank the reviewer for making us aware of this distinction. After reviewing this distinction of capacity and capability, we’ve decided to reword the above-mentioned phrase to: 'Lack of expertise and internal capability’, and the thematic category to: 'Factors related to technical capability and evaluation skills’. Throughout the paper (and Table 1-3), we’ve applied this distinction more carefully, particularly in the discussion and conclusion. Additionally, we’ve addressed the distinction explicitly in ‘Implications of findings’ section. It now reads:
“The thematic synthesis demonstrated that the main barriers for TSOs to undertake evaluation was related to organisational capacity and capability. One may think of capacity as ‘how much’ an organisation can do, relating to issues around resources, and capability as ‘how well’ an organisation can operate, thus relating to the skills, knowledge and confidence of organisations [65, 66]. These issues were reflected in the most reported categories of factors operating as barriers; ‘factors related to lack of resources’ (capacity), and ‘factors related to technical capability and evaluation skills’ (capability).

Importantly, these findings suggest that many TSOs are not receiving the appropriate capacity or capability support in evaluating their services. The significance of these barriers is further illustrated by the most reported categories of facilitating factors being to receive the appropriate support to undertake evaluation. The prevalence of these categories implies a rather unfortunate situation in which TSOs are faced with a growing pressure to evidence their performance to secure grants or contracts, but without being subject to the necessary support to undertake meaningful evaluation.” (p17-18, lines 430-446)

2. Provide more information on the assessment of study quality

It is not clear on what basis studies were assessed as being of 'high', 'medium' or 'low' quality (p.10, line 244). The checklists in Appx A and B are helpful, but it would be helpful to say a little more about this in the text (see also p.13, lines 309-316, where reference is made to 'achieving a score' - what does this mean?).

We’ve clarified the language regarding the quality scores and elaborated on how we decided on the quality criteria.

Regarding ‘achieving a score’:

“Three studies (12.5%) were deemed to be of ‘high quality’ [32, 51, 64], fifteen (62.5%) of ‘medium quality’ [27, 29, 44–48, 50, 52, 53, 56, 57, 59–61], and the remaining six studies (25%) [28, 49, 55, 58, 62, 63] were assessed to be of ‘low quality’.” (p13, lines 323-327)
We’ve also added a short elaboration on how the quality assessment was done:

“The basis of the quality ranking was informed by the number of checklist items fulfilled by the individual studies, but the final assessment of quality was based on an overall judgement of the value of the individual findings, considering the methodology of the included studies.” (p11, lines 255-258)

3. Modify the arguments supporting the promotion of EBP in the discussion and concluding sections

The rationale for promoting EBP seems a little overplayed and, I thought, stretched beyond the results of the review. I felt that the final paragraph on p.18 (and through to p.19) involves something of an unwarranted translation: a reported problem with indicators ('Challenges in identifying accepted outcome and impact indicators') is consensus procedures from EBP remain largely unused by the third sector…' (p.19, line 485-6)? This reads like an over-claim, and is either in need of a supportive reference or argument, or, I would suggest, should be toned down. Finally, there is a need for care to avoid slippage in the reference to '….may reinforce poor evaluation practice among TSOs' (p.19, line 472). The review does not provide evidence of poor evaluation practice, only of reported and interpreted barriers and facilitators.

We thank the reviewer for this comment. We’ve clarified that using stakeholder inclusive methods from EBP should not only be seen a tool to construct new evaluation criteria, but also as a mean to identify stakeholders needs.

Further, we’ve toned down our language on the promise of EBP, so that it should be clearly presented that whether EBP consensus procedures are the right way forward should be considered an empirical question, which should be tested in practice.

These changes appear in the ‘Ways forward’ (p19-21, lines 484-524) and in the ‘Conclusions’ (p23-24, lines 582-609) sections.
4. Necessary and sufficient conditions

The paper concludes with the argument that ‘…it is central not to focus on individual barriers (such as financial resources), but to reach consensus, or at least have a procedure for consensus, to determine what evaluation criteria different types of TSOs require…’ (p.22, lines 544-555). The aim of the argument here seems to be to draw attention away from financial barriers and towards evaluation criteria. This does not seem to be a justifiable step when expressed in these terms. For example, financial barriers seemed actually to win out in the 'vote count' against 'identifying accepted outcome and impact indicators' (14 against 13). Rather than overlook financial barriers, as the present wording of the text suggests (e.g. 'it is central not to focus on…'), my suggestion would be to distinguish between 'necessary' and 'sufficient' conditions for improved evaluation practice in the third sector. My reading of the systematic review is that overcoming financial and similar 'capacity' barriers would be necessary, but not sufficient, to improve practice (given the existence/prevalence of other factors), but, equally, securing a consensus on evaluation criteria (and other 'capability' issues) is also necessary, but not sufficient (given the significance of financial and other capacity constraints). Hence the review overall might conclude that 'capacity' and 'capability' (in this case in evaluation practice) are separate but very closely related.

We thank the reviewer for this comment. We’ve rephrased our ‘Ways forward’ (p19-21, lines 484-524) and ‘Conclusions’ (p23-24, line 582-609) sections accordingly. Specifically, we’ve changed our wording so that it hopefully becomes clear that in order to meet the identified barriers and to enhance facilitators, it is important to consider a full picture of factors influencing evaluation, which the lens of necessary and sufficient conditions allows for.

Minor notes, queries, typos, etc.

We thank the reviewer for the very rigorous and helpful comments/suggestions below.

-Title: do you really 'experience' a facilitator? Can you think of another word, e.g. 'encounter'? or just 'What are the barriers and facilitators for good evaluation practice in third sector organisations? A systematic review'

We have changed the title to: “What are the barriers and facilitators for third sector organisations (non-profits) to evaluate their services? A systematic review”. We thought that ‘good evaluation
practice’ would be misleading, as we’re not assessing the quality of the evaluation practice covered in the included studies.

-p.2, lines 31-32 (abstract): is it true to say that 'many TSOs fail to evaluate their activities following the principles of evidence based practice?' how do we know? The review is of reported barriers and facilitators, not of actual evaluation practice.

We thank the reviewer for this comment. This sentence was meant to set the background for why it may be relevant to investigate barriers and facilitators, i.e. considering the evidence that many TSOs struggle to evaluate, it is important to investigate why that might be the case. However, we have changed the language to be less absolute:

“However, in responding to the increased pressure to undertake evaluation, research suggests that many TSOs struggle to evaluate their activities following the principles of evidence-based practice (EBP).” (p2, lines 31-33)

-p.4, lines 82-83: the cited definition of EBP would seem to be irrelevant to many TSOs, given the reference to 'clinical expertise' and 'patients'. I'd suggest using a broader or more generic definition.

In response to this comment and comment 1 by Reviewer 1, we have removed the quote, as we can see that it may be misleading in this case. Instead we have framed the sentence to be about the importance of evaluation:

“However, despite the increasing emphasis on impact and performance evaluation by government and commissioners, many third sector organisations (TSOs) fail to evaluate their activities following the principles of evidence-based practice (EBP), which mandate the adherence to rigorous, reproducible and systematic methods [9]. While it is generally acknowledged that monitoring and evaluation according to best practice is important in the delivery of services, there has been little systematic effort to investigate why the third sector is struggling to evidence its impact, despite becoming a growing provider of social and health services [10–13].“ (p4, para 1)
We thank the reviewer for this comment and have amended the text as suggested.

This definition of TSOs was generated by the Johns Hopkins Comparative Nonprofit Sector project, so perhaps the reference should be to the original. In a UK setting, the reference is Kendall, J. and Knapp, M. (1996) The Voluntary Sector in the UK (Manchester, Manchester University Press).

We thank the reviewer for this comment. It is true that the definition is derived from John Hopkins Comparative Nonprofit Sector project, and we have now referenced it accordingly. However, we have kept the quote by Hardwick et al 2015, as we find it a concise summary of the definition.

The sentence now reads:

“In the context of this review, we will use the broader term ‘third sector organisation’ (TSO) and follow the structural/operational definition [16], which can be summarised as “…organisations which are formally organised; non-profit distributing; constitutionally independent from the state; self-governing and benefiting from some form of voluntarism (e.g. with volunteer (unpaid) Trustees or Board members or using volunteers in the delivery of services).” [17].” (p4, lines 95-100)

I don't think it is true that 'most' TSOs in the UK focus on delivering 'social services' Does the Almanac data really show this?

The almanac (https://data.ncvo.org.uk/a/almanac17/size-and-scope-2/) shows that the biggest (both in terms of number of TSOs and spending) activity type of UK TSOs are social services. However, we realise that ‘most’ might be slightly misleading and we’ve rephrased ‘most’ to ‘many’.
-p.5, line 103: 'annual spending of £45.5bn...' should be 'annual income (spending was £43.3bn)'

We thank the reviewer for this comment and have amended the text as suggested.

-p.5, lines 104-106: 'Of these budgets...' is confusing, as it points attention to the two figures in the preceding sentence. But the government's £15.3bn funding to the sector is (a) greater than the two figures in the preceding sentence, and thus (b) should relate to the overall sector income figure, of £45.5bn

We have changed the sentence to:

“Of its annual income (£45.5 bn 2014/2015), it has been estimated that the government provides £15.3 (33.3%) billion to the third sector, largely through contract-based commissioning.” (p5, lines 110-112)

-p.5, line 112: after 'their sample...' suggest adding in what the sample figure was, i.e. 8% of what?

We thank the reviewer for this comment and have amended the text as suggested.

It now reads:

“For example, a 2010 report developed by the Charity Finance Group (CFG) found that only 8% of their sample of 75 TSOs provided external evidence on their impact.” (p5, lines 117-119)

-p.5, line 112: a reference is given here to 'Breckell et al 2010', but this seems out of line with the referencing approach taken in the rest of the paper

We thank the reviewer for this comment and have amended the text as suggested.
We thank the reviewer for this comment and have amended the text as suggested.

I'd suggest rephrasing the passage here, as currently it sounds like it was remiss of Hardwick et al not to have considered evaluation processes, rather than simply that this was not their chosen focus.

We thank the reviewer for this comment. We rephrased the passage to:

“However, the focus of the review was on the knowledge mobilisation of third sector practitioners, and not on the evaluation process [17].”

Also, we’ve deleted the following sentence: “As a result of this focus, the subsequent search strategy was predominantly designed to identify studies concerned with how TSOs use and mobilise evidence, meaning that research on the barriers and facilitators related to the evaluation process was not addressed”

We’ve included the following sentence: “There were no date restrictions for studies to be included in the review.”

We thank the reviewer for this comment and have amended the text as suggested.
-p.9, line 219-220: the statement '….studies were either qualitative, mixed methods or cross-sectional…' does not seem to make sense, as the three categories here are not mutually exclusive. Quantitative research can be either longitudinal or cross-sectional, but so can qualitative research. The paper seems to imply that 'cross-sectional' studies are automatically quantitative (see also p.10, line 235). Maybe it would be better to say 'qualitative, survey-based or mixed methods…'

We thank the reviewer for this comment and have amended the text as suggested. The sentence now reads:

“The focus of this review is on factors identified by third sector practitioners and included studies were thus either qualitative, survey-based or mixed methods” (p10, lines 229-231)

-p.10, line 234: It isn't clear why the variability of numbers of TSOs in the studies is a reason for undertaking a quality appraisal. I'm not sure a reason is needed - assessing quality is an important step in its own right.

We've deleted the sentence: “Due to the variability of the number of TSOs covered within studies, all studies were appraised using relevant critical appraisal criteria [34].”

-p.10, line 253: five studies were excluded because they were '…about community participatory research processes…' It isn't clear what this means and why they were excluded - an additional explanation is needed.

We rephrased this group of excluded studies to: “…focusing on the collaborations between researchers and community programmes…” (p11, lines 266-268)

-p.11, line 265: full text needed for 'LMIC' before acronym

We rephrased the sentence to: “Only one study (4.2%) was conducted entirely in a low and middle-income country (LMIC), and one study (4.2%) was conducted in multiple contexts - US, Columbia, Guam and Puerto Rico.” (p11-12, lines 279-281)
p.11, line 270: Suggest sub-title as 'Samples of TSOs in included studies'

We thank the reviewer for this comment and have amended the text and Figure 2 as suggested.

p.11, line 275: '…a 1000 or more TSOs…' rather than '…more than 1000 TSOs…'

We thank the reviewer for this comment and have amended the text as suggested.

p.12, line 282: the bias towards 'larger organisations' needs an explanation or definition, since 'large' and 'small' in the context of TSOs are used vaguely and inconsistently - are the income bands used in the subsequent sentence what you mean here?

We thank the reviewer for this comment and have amended so that it should be clear that the bias is self-reported. Unfortunately, and as pointed out by the reviewer, many studies in the literature do not specify what is meant by 'larger organisations'.

“Seventeen of the studies (70.8%) included a mixed population TSOs in terms of income [27–29, 44, 45, 47, 48, 50, 52–55, 57–61], of which eight had a self-reported bias towards larger organisations [27, 28, 45, 48, 53, 58, 60, 61].” (p12, lines 295-297)

p.12, line 283: add 'annual income' before '….below £100,000'

We thank the reviewer for this comment and have amended the text as suggested.

p.12, lines 284, 285: '£1million' not '1£million'

We thank the reviewer for this comment and have amended the text as suggested.
-p.12, line 291: 'in' rather than 'with' social and human services?

We thank the reviewer for this comment and have amended the text as suggested.

-p.13, line 314-315: suggest adding 'surveys' before '…of a cross-sectional nature…' and then remove 'survey' from '…validate the survey findings…'

We thank the reviewer for this comment and have amended the text as suggested.

-p.13, line 329: 'appropriate evaluation goals (15/24)' should be '…(16/24)…' if table 2 is correct.

We thank the reviewer for this comment and have amended the text as suggested.

-p.14, line 335: is table 1 really needed, given that it is fully described in the preceding text? I'd suggest one or the other is needed but not both.

We thank the reviewer for this comment. Whilst we recognise that both tables entail the same information, it is common practice in this type of reviews to summarise the descriptive results by including a short table that presents the most reported factors.

-p.14, line 340: suggest removing 'survey' before 'respondents', as the findings in some of the studies do not come from surveys

We thank the reviewer for this comment and have amended the text as suggested.

-p.14, lines 343-344: 'supports' and 'requires', since 'management' is singular
We thank the reviewer for this comment and have amended the text as suggested.

-p.15, line 372: '...were those...' rather than '...was factors...'

We thank the reviewer for this comment and have amended the text as suggested.

-p.15, lines 380-382: I found this sentence to be very unclear: can it be reworded?

We thank the reviewer for this comment and have amended the text as suggested. It not reads: “Table 3 and 4 illustrate that the salience of themes corresponded to how representative the themes were of the included studies. However, ‘organisational culture and management’ was an exception, in that it represented more of the included studies, compared to other top categories of facilitators.” (p16, lines 395-399)

-p.16, line 384: suggest '...only two studies, both qualitative...' 

We thank the reviewer for this comment and have amended the text as suggested.

-p.16, line 391: because it isn't immediately obvious what a 'sensitivity analysis' entails, I'd suggest adding 'by' between 'conducted' and 'including'

We thank the reviewer for this comment and have amended the text as suggested.

-p.17, line 415: suggest 'for example' rather than 'e.g.' here (and in subsequent places)

We thank the reviewer for this comment and have amended the text as suggested (also in the instances that we use ‘e.g.’ earlier in the paper).
We thank the reviewer for this comment and have amended the text as suggested.

-p.17, line 416: '…appropriate support…' rather than '…the appropriate support…'

We thank the reviewer for this comment and have amended the text as suggested.

-p.17, line 418: I'm not sure what a 'commissioning contract' is. I'd suggest a more general reference - 'grants or contracts'

We thank the reviewer for this comment and have amended the text as suggested.

-p.17, line 422: 'procedures'

We thank the reviewer for this comment and have amended the text as suggested.

-p.17, line 430: suggest adding 'and facilitators' after 'external barriers', since the preceding paragraph is about both barriers and facilitators

We thank the reviewer for this comment and have amended the text as suggested.

-p.17, line 432: move 'both' to before 'barriers'

We thank the reviewer for this comment and have amended the text as suggested.

-p.18, line 456: suggest 'The most cited group of factors…' rather than 'One of the most cited themes of factors…'

We thank the reviewer for this comment and have amended the text as suggested.
-p.18: is it worth a line or two exploring why 'lack of resources' features as a barrier, but 'more resources' doesn't feature so much as a facilitator?

We thank the reviewer for this comment and have tried to elaborate on this by rephrasing the first paragraph in the ‘Ways forward’ section. Please see below comment for details.

-p.18, line 459: it would be important to recognise that 'lack of resources' was cited by just as many times as lack of evaluation skills.

We thank the reviewer for this comment and have rephrased the paragraph. It now reads:

“The most cited groups of factors facilitating TSOs to undertake evaluation was getting the appropriate support to evaluate (n=19), in which ‘financial support’ was only identified twice as a facilitating factor under this theme. The low reporting of financial resources as a facilitating factor is noteworthy, considering that ‘lack of resources’ (n=36) was among the most cited categories of factors operating as barrier, along with the lack of evaluation skills and capability (n=36). This suggest that increasing financial support for TSOs to evaluate might be a necessary factor to improve evaluation practice, but that it is not sufficient in isolation. Rather, the analysis demonstrates that TSOs need improved support in the entire evaluation process, which would require attention to both capacity (more resources) and capability (e.g. staff training and collaboration with experts and researchers).” (p19-20, lines 485-595)

-p.19, line 466: 'social impact bonds' are not evaluation tools or frameworks; they are financial mechanisms. Also it should be 'Social Return on Investment' not '…of investment'

We thank the reviewer for this comment and have amended the text as suggested.

-p.19, line 469: typo: '…no any…'

We thank the reviewer for this comment and have amended the text as suggested.
p.20, line 509: suggest using 'fields' or 'types of organisation' rather than 'sectors' here, since 'sector' is also used typically as a label for the aggregate of voluntary, community and non-profit organisations

We thank the reviewer for this comment and have amended the text as suggested.

p.21, lines 522-523: '…most third sector spending originates from Western TSOs…' Is this true? A reference is needed to support this claim.


“However, given that third sector spending tends to be higher in Western TSOs, these findings may be generalisable to a larger population of organisations.” (p22, lines 561-563)

p.21, line 533: You might also add Jenny Harlock's TSRC Working Paper from June 2014 (No.123) to reference 87, as it reports commissioners’ views on evaluation requirements.

We thank the reviewer for this comment and have amended the text as suggested.

p.22, line 552: '….are put at risk….' (not 'is')

We thank the reviewer for this comment and have amended the text as suggested.
-p.22, line 559: suggest '…to assess with confidence…' rather than 'certainty', which seems like an impossibly high threshold.

We thank the reviewer for this comment and have amended the text as suggested.

-pp.24-32 References: some of the references are not fully cited. For example, TSRC working papers have details missing: e.g. Reference No.4 (page 24) would be 'Harlock, J. (2013) 'Impact measurement practice in the UK third sector: a review of emerging evidence' TSRC Working Paper 106, July 2013 (Birmingham, University of Birmingham), followed by a direct weblink?; References 28, 32 and 33 seem to be of the same report?; References 1 and 75 are to the same item.

We have amended the references, so that they should appear as suggested.