

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Identifying positively deviant elderly medical wards using routinely collected NHS Safety Thermometer data: an observational study.
AUTHORS	Baxter, Ruth; Taylor, Natalie; Kellar, Ian; Pye, Victoria; Mohammed, Mohammed; Lawton, Rebecca

VERSION 1 – REVIEW

REVIEWER	Jodie Gary Texas A&M University College of Nursing I have a research interest in the research of healthcare outcomes through the lens of positive deviance. I am on the other hand very familiar with the current positive deviance literature as well as other work by the current authors. I do not feel that I have any current competing interests. It is a pleasure to review this content.
REVIEW RETURNED	30-Oct-2017

GENERAL COMMENTS	<p>This is a VERY timely topic to those interested in positive deviance (PD) as a methodology for impacting healthcare. This study addresses the a specific gap in PD literature and provides readers with a description of how the methodology can be replicated as well as limitations to be considered. PD needs studies such as this one with data to support PD as a viable research method. Strengths and limitations are adequately addressed and evaluated. The discussion of the method was dissected with reference to issues known in PD literature; this adds to the interest of the paper for the reader. The realistic discussion concerning the use of the methodology is appreciated.</p> <p>Here are some minor considerations:</p> <ol style="list-style-type: none">1. p. 5 line 10 - consider removing the word "quite"2. Ensure that all acronyms are defined with first use. example is NHS noted in the abstract on on page 5 line 39, but then defined on p. 63. p. 6, line 22 - consider rewording the phrase "will be published elsewhere"4. Under Analysis, p. 7, consider a brief descriptor of SPC. I am familiar with it, but this may be helpful to the readers5. p. 14 line 52 - consider using "do not" <p>This has been a pleasure to review and the work in this area is greatly appreciated.</p>
-------------------------	--

REVIEWER	Kathleen M. Sutcliffe Johns Hopkins University United States of America
REVIEW RETURNED	03-Nov-2017

GENERAL COMMENTS	<p>The goal of the study is to develop a pragmatic replicable method to identify 'positively deviant' medical wards. Specifically, the manuscript reports findings from a two-phased study seeking to identify positively deviant, exceptionally high performing, elderly medical wards within a region of northern England. During phase one the authors drew on National Health Service (NHS) Safety Thermometer data to identify a set of exceptionally safe wards in 13 acute NHS trusts. Concomitantly the authors identify a set of matched comparison wards demonstrating above average safety performance. In the second phase, the authors surveyed patients and staff perceptions of safety using indices that have previously been validated. The goal was to determine if staff and patients' perceptions were congruent with phase 1 findings. Although there were some instances of disagreement, for the most part the findings revealed five high performing (safe, perhaps positively deviant) wards, which seemed to perform better than matched counterparts. The authors suggest that they wish to advance theory in the domain of positive deviance, and from a practical standpoint to better understand how safe care is enabled and enacted in the positively deviant wards (in contrast to the average wards), and to explore the extent to which these existing strategies/behaviors are sustainable. The paper is tightly written. The methods seem to be quite well thought out and rigorous. The work to date seems to be well executed. I had a couple of questions. For example, it was a bit difficult to determine exactly how the perceptual data were used. I understand they were averaged. I was curious about the extent to which people's perceptions with the wards were in agreement. Is there a way to conduct an interrater reliability/agreement statistic? It might give some insight about what is going on with these perceptual data.</p> <p>The research is important and timely. I urge the authors to think harder about some of what they are taking for granted (e.g., re-look at some of the statements in the discussion). For example, there seems to be an assumption that what is going on in the positive deviant wards (i.e. what is keeping patients safe) can be diffused, implemented, or in any way possible, generalized to other contexts. This assumption is questionable. The organizational learning literature has questioned this kind of idea and shown in several studies that transferring practices/routines from one context to the next with the same results is almost impossible. Perhaps more important, however, is whether participants of the high performing wards would even be able to articulate or discern with any accuracy, validity or reliability why/how things are working so well. It is hard to know all the things that are keeping a system safe at one moment in time. Additionally, organizational contexts are always unique—we cannot assume that any unit or organization – positively deviant or not – face the same constraints as others. Routines/actions are always contextualized, or should be. If they aren't they are likely to be less effective.</p> <p>I appreciated the opportunity to read this paper. I wish the authors luck with this important work.</p>
REVIEWER	Kelly Smith
REVIEW RETURNED	MedStar Institute for Quality and Safety, United States of America
REVIEW RETURNED	14-Nov-2017
GENERAL COMMENTS	This was a well written and well researched paper. The approach is innovative and provides a clear description that could be replicated

	<p>by others interested in applying the methodology of positive deviance to patient safety improvements. The integration of staff level and patient level data is an asset, although the small sample sizes limit the ability to do more sophisticated analyses. Given the small sample sizes, focus groups may have yielded more rich evidence and overcome some of the limitations described by the authors.</p> <p>One typographical error on page 8, line 49 "that" should be "than".</p> <p>Overall, this is a novel paper describing a reproducible approach to identifying positively deviant wards/units. It is a valuable contribution to our field.</p>
--	---

REVIEWER	Tamar Klaiman AccessMatters USA
REVIEW RETURNED	14-Nov-2017

GENERAL COMMENTS	<p>The goal of this article was to develop and critically appraise a robust yet pragmatic method for identifying positive deviants at ward level using a routinely collected, broad outcome of safety. Overall, I think the manuscript offers a contribution to the field. Some suggestions for clarity and improvement are noted below:</p> <ul style="list-style-type: none"> - The data collection tools utilized to identify the positive deviants appear to only include outcome measures, which is fine, but I am not clear what the patient and staff survey add to the study. There is evidence that patient perspectives about satisfaction often do not align with medical quality and improved outcomes (ie. nice doctors aren't always the best doctors). I also don't see any metrics related to process which is where the rubber meets the road in terms of actionable change to improve outcomes. Perhaps there are more data that elucidate these in your next manuscript, but it's not clear to me that what you have found can make a practical impact without additional metrics (including qualitative data). - In Table 1, it would be helpful to also include data on negative deviants. You address why negative deviants weren't included, but the PD and comparison groups are only very slightly varied. It would be useful to see the negative deviants compared to both PDs and average wards. That would also make a stronger case for using this method to identify high performers. - Some of the tables and graphs are hard to decipher without color – make sure they will be printed in color or change them to grayscale. - Table 2 is confusing. It appears that some wards were identified as PDs but performed poorly in some areas. Can you explain why they remain PDs? I would encourage you to focus on the ST results and dig into why these wards perform better. I think adding the other metrics, particularly those based on perception of patients and friends and family, add a layer of complications that isn't necessary. - In the discussion, you noted that the PDs performed well over time, but you only have a year's worth of data.
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Comment: p. 5 line 10 - consider removing the word "quite"

Response: The word 'quite' has been removed to reduce ambiguity.

Comment: p. 6, line 22 - consider rewording the phrase "will be published elsewhere"

Response: Phrase reworded to say: "The qualitative findings from this wider application (stage 2 of the framework) will be published separately".

Comment: Under Analysis, p. 7, consider a brief descriptor of SPC. I am familiar with it, but this may be helpful to the readers

Response: A brief descriptor of SPC has been included:

"Statistical Process Control (SPC) methods such as run charts and funnel plots are increasingly being promoted to assess variation within healthcare.(40, 41) Performance variations exist within any stable system and SPC methods can be used to distinguish between variation that occurs by chance (i.e. noise in the system) and variation that has an assignable cause (i.e. a signal of positive deviance).(42)"

Comment: p. 14 line 52 - consider using "do not"

Response: The sentence has been reworded.

Reviewer 2

Comment: It was a bit difficult to determine exactly how the perceptual data were used. I understand they were averaged. I was curious about the extent to which people's perceptions with the wards were in agreement. Is there a way to conduct an interrater reliability/agreement statistic? It might give some insight about what is going on with these perceptual data.

Response: To clarify how the data from the staff and patient surveys were collected and analysed we have included more detail in the Phase 2 procedure and analysis sections of the paper.

Regarding the extent to which individual staff and patient perceptions of safety varied around the mean scores, Table 1 does include the standard deviations for each average score. To ensure clarity, we have defined 'standard deviation' within Table 1 as this was previously just referred to as 'SD'. We have also added a table into supplementary file 4 that contains the minimum and maximum values for each measure to provide additional information about the extent to which staff and patient perceptions varied within wards.

Patient and staff perceptions of safety were measured using different survey tools (the PMOS for patients and a measure of safety culture for staff), therefore, we do not believe that it is appropriate to calculate an agreement statistic.

Comment: I urge the authors to think harder about some of what they are taking for granted (e.g., re-look at some of the statements in the discussion).

- For example, there seems to be an assumption that what is going on in the positive deviant wards (i.e. what is keeping patients safe) can be diffused, implemented, or in any way possible, generalized to other contexts. This assumption is questionable.
- Perhaps more important, however, is whether participants of the high performing wards would even be able to articulate or discern with any accuracy, validity or reliability why/how things are working so well.
- Additionally, organizational contexts are always unique—we cannot assume that any unit or organization – positively deviant or not – face the same constraints as others.

Response: The reviewer has highlighted a number of challenges and assumptions that exist when applying the positive deviance approach. From our perspective, the first two assumptions that have been listed relate specifically to stage 2 and 4 of the Bradley framework. We highlight at the top of page 7 that this study focuses solely on identifying positive deviants (stage 1 of the framework) and so we have not included an extensive discussion of these assumptions so as not to confuse the message and focus of the paper. A more robust discussion of these assumptions will be included in our future publications; however, due to their importance we have made reference to them on page

15 of this paper. We have also made some small changes throughout the manuscript to reemphasise the focus of this paper and to clarify these assumptions where possible.

We consider the reviewer's final point to be one of the key challenges to identifying positive deviants and the manuscript does contain a brief discussion of this on page 15/16 (ensuring homogeneity across the population). However, due to its importance, we have added more detail to expand and clarify this aspect of the discussion.

Reviewer 3

Comment: One typographical error on page 8, line 49 "that" should be "than".

Response: Thank you for pointing this out - the typographical error has been corrected.

Reviewer 4

Comment: The data collection tools utilized to identify the positive deviants appear to only include outcome measures, which is fine, but I am not clear what the patient and staff survey add to the study. There is evidence that patient perspectives about satisfaction often do not align with medical quality and improved outcomes (ie. nice doctors aren't always the best doctors). I also don't see any metrics related to process which is where the rubber meets the road in terms of actionable change to improve outcomes. Perhaps there are more data that elucidate these in your next manuscript, but it's not clear to me that what you have found can make a practical impact without additional metrics (including qualitative data).

Response: We hope to have provided greater clarity on the role of the patient and staff surveys within the study by more explicitly defining and presenting the study aims within the introduction section and by restated the aims at the very beginning of the discussion section.

The reviewer is correct in thinking that our next manuscript will explore the actionable changes that can be made to improve outcomes (i.e. stage 2 of the Bradley framework). Our forthcoming paper will present qualitative data that have been collected to explore and generate hypotheses about how the positively deviant ward teams deliver exceptionally safe patient care.

Comment: In Table 1, it would be helpful to also include data on negative deviants. You address why negative deviants weren't included, but the PD and comparison groups are only very slightly varied. It would be useful to see the negative deviants compared to both PDs and average wards. That would also make a stronger case for using this method to identify high performers.

Response: Unfortunately, the PMOS, FFT, CQUIN, and PSG measures are not routinely collected and/or published at ward level within the NHS. These data had to be collected by the researcher during phase 2 of the study and, due to resource constraints, it was only possible to do this on the positively deviant and comparison wards. As a result we are unable to include the data for negatively deviant wards in Table 1.

However, the results section of Phase 1 (page 8) reports the range in NHS ST 'harmfree care' performance for all wards in the region. This highlights how the most negatively deviant and most positively deviant wards in the region compare. A comparison of positive deviants with negative deviants is also available through Figure 1 and Supplementary File 1 which plot and tabulate the ST performances of all wards in the region.

Comment: Some of the tables and graphs are hard to decipher without color – make sure they will be printed in color or change them to grayscale.

Response: Where possible figures have been presented in colour. Table 2 and supplementary files will be printed in colour.

Comment: Table 2 is confusing. It appears that some wards were identified as PDs but performed poorly in some areas. Can you explain why they remain PDs? I would encourage you to focus on the ST results and dig into why these wards perform better. I think adding the other metrics, particularly those based on perception of patients and friends and family, add a layer of complications that isn't necessary.

Response: The aim of Phase 2 of this study was to explore whether staff and patient perceptions of safety were similar to, or at odds with, the routinely collected ST data that had been used to identify the positively deviant wards during Phase 1 (i.e. whether staff and patients corroborated the identification of positive deviant). Table 2 seeks to answer this question by presenting how the positively deviant wards (as identified by ST data) performed in terms of patient and staff perceptions of safety. To clarify the information that is presented, we have changed the title of Table 2 ("A visual representation of how positively deviant and comparison wards, that were identified using routinely collected ST data, ranked on patient and staff perceptions of safety") and more detail has been added into the column headings and footnotes. Furthermore, by stating the aims of the study more explicitly within the introduction section, we hope that the purpose of this table and results section will be clearer.

This study forms part of a wider application of the positive deviance approach (see page 6). As part of this wider application, we have collected qualitative data to explore why the positively deviant wards perform better than the comparator wards. This paper focuses on critiquing the methods that can be used to identify positive deviants (stage 1 of the approach) and we will seek to publish the qualitative findings (which focus on stage 2 of the approach) separately.

Comment: In the discussion, you noted that the PDs performed well over time, but you only have a year's worth of data.

Response: The sentence has been changed to add clarity and reduce ambiguity:

"... identifying a distinct and statistically different group of wards that not only ranked best within the region but also outperformed their NHS Trust (organisation), performed consistently over time (12 months), and"

The ST data is routinely collected on a monthly basis. Data were analysed over a 12 month period to identify positive deviants and run charts were created to ensure that their exceptional performance was not just a 'one off' success.

VERSION 2 – REVIEW

REVIEWER	Jodie Gary Texas A&M University United States
REVIEW RETURNED	19-Dec-2017

GENERAL COMMENTS	Thank you for the opportunity to follow-up on the revisions of this manuscript. The revisions have provided clarity to the manuscript.
-------------------------	--

REVIEWER	Kathleen Sutcliffe Johns Hopkins University United States of America
REVIEW RETURNED	19-Dec-2017

GENERAL COMMENTS	The revised paper includes elements related to issues and concerns
-------------------------	--

	I raised in my previous review. I appreciate the work the authors did to address these concerns. I appreciated the opportunity to read this paper. I wish the authors luck with this important work.
--	---

REVIEWER	Tamar Klaiman AccessMatters USA
REVIEW RETURNED	11-Dec-2017

GENERAL COMMENTS	The changes made satisfy my concerns. Please be sure to refer to data in the plural.
-------------------------	--