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)	The Effectiveness of Policy and Risk Targeting for Opioid-Related Risk Mitigation: A Randomized Program Evaluation with Stepped-Wedge Design
AUTHORS	Minegishi, Taeko; Garrido, Melissa; Pizer, Steven; Frakt, Austin

VERSION 1 – REVIEW

REVIEWER	Steven K. Dobscha MD Oregon Health & Sciences University VA Portland Health Care System, Oregon USA
REVIEW RETURNED	05-Nov-2017

GENERAL COMMENTS	The purpose of this manuscript is to describe the study protocol for a randomized program evaluation of the Stratification Tool for Opioid Risk Mitigation (STORM). The manuscript is very clearly written and complete, efficient, and would be of great interest to pain and opioid researchers as well as others planning program evaluations or studies employing stepped wedge designs. It is a very interesting and innovative project. I have only a few comments/questions: 1. The authors note that VHA Central Office is going to release a policy memo in Fall 2017. To my knowledge this has not yet happened. In light of the ongoing and sometimes unpredictable political winds, it might be wise to reframe this as the policy memo is anticipated to come out in the near future. I would also recommend that Figure 2 which uses quotations to describe the policy be redone to not include quotations but instead key messages probably more in bulleted form that are expected in the policy, just in case there are changes in wording 2. Related to above, I do wonder if the investigators have a contingency plan for conducting this study or elements of this study should this policy memo never come out. If they do, they might wish to consider describing it in this manuscript if they feel it is appropriate. 3. I have some concerns about the additional oversight and facilitation to be provided by OMHSP. First, I would just like to know a bit more about what that might entail since it will affect facilities being studied. Second, since this facilitation is very likely to be staged over time (ie OMHSP will not be able to provide oversight/facilitation to all indicated facilities at the same time) and likely to vary in intensity, it would be important to try to include these factors in the models as they could affect study outcomes (they would be expected to)—I don't see now that these factors are included in the models 4. In the recruitment section on page 6, it would be nice to see an estimate of numbers of Veterans in the overall analytic cohort—in

	the top 10% of risk scores. See lines 22-23.

REVIEWER	Michael Saenger Atlanta VA Health Care System, USA; Emory University, USA
REVIEW RETURNED	10-Nov-2017

GENERAL COMMENTS	Bravo for asking and trying to answer these two questions, and for helping to raise the bar in care and in policy implementation.

REVIEWER	André Scherag Institute of Medical Statistics, Computer Sciences and Documentation Jena University Hospital, Jena, Germany
REVIEW RETURNED	10-Jan-2018

GENERAL COMMENTS	I was asked to focus on the methods/statistics of this proposal. It is positive that the authors plan a randomized comparison to answer their two main research questions. Question 1 is relatively clear and the design matches to the research question. However, research question 2 and the operationalization differ - if the change of the included top is the aim, then this should be mirrored in the research question. Again, this also has an impact on the design and has to be reflected in both the sample size estimation (which should focus on the unit of randomization which are the VHA medical centers) and the analysis (which has to deal with both the center effects and the temporal pattern). What if interactions (between the experimental groups) exist? Moreover, the authors should comment on why they suggest using SAE rates (I guess some kind of cumulative incidence?) in the sample size estimation part and time-to-event analyses in the methods? That will likely confuse readers. Finally, I would also suggest reading CONSORT extensions for cluster randomized studies in order to be able to address their special reporting requirements.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Steven K. Dobscha MD

Institution and Country: Oregon Health & Sciences University, VA Portland Health Care System, Oregon, USA

The purpose of this manuscript is to describe the study protocol for a randomized program evaluation of the Stratification Tool for Opioid Risk Mitigation (STORM). The manuscript is very clearly written and complete, efficient, and would be of great interest to pain and opioid researchers as well as others planning program evaluations or studies employing stepped wedge designs. It is a very interesting and innovative project. I have only a few comments/questions:

1. The authors note that VHA Central Office is going to release a policy memo in Fall 2017. To my knowledge this has not yet happened. In light of the ongoing and sometimes unpredictable political winds, it might be wise to reframe this as the policy memo is anticipated to come out in the near future. I would also recommend that Figure 2 which uses quotations to describe the policy be redone to not include quotations but instead key messages probably more in bulleted form that are expected in the policy, just in case there are changes in wording.

- Yes, we couldn't agree more. When we were writing this manuscript, we were optimistic that the memo would be released in Fall 2017. We modified the sentence on page 3 of the revised manuscript to say "In the near future".

- We agree with the reviewer's comment on Figure 2. We have modified Figure 2 into a Table and summarized the key policy differences between the treatment and the control group.

2. Related to above, I do wonder if the investigators have a contingency plan for conducting this study or elements of this study should this policy memo never come out. If they do, they might wish to consider describing it in this manuscript if they feel it is appropriate.

- Since the study is designed to test a VHA policy, we do not have a contingency plan. If the memo does not get released, we will not have a study. However, we have been in close contact with leadership at the VHA operations office responsible for the memo, and we have been assured that it is in the final stages of preparation before release. Thus, we are very confident that the memo will be released in the near future.

3. I have some concerns about the additional oversight and facilitation to be provided by OMHSP. First, I would just like to know a bit more about what that might entail since it will affect facilities being studied. Second, since this facilitation is very likely to be staged over time (ie OMHSP will not be able to provide oversight/facilitation to all indicated facilities at the same time) and likely to vary in intensity, it would be important to try to include these factors in the models as they could affect study outcomes (they would be expected to)—I don't see now that these factors are included in the models

- Thank you for highlighting these important issues.

- Medical centers in the treatment group that do not meet the targeted metric of > 97% review of very high risk patients may receive oversight and facilitation, which includes technical assistance and action planning as well as 1) addition of the metric (i.e. >97% review of very high risk patients) to the facility's existing improvement goals and 2) quarterly reports from facility point(s) of contact detailing progress toward executing an action plan to meet the metric. We have clarified these details in Table 1 and on page 4 of the manuscript.

- The oversight and facilitation during the study period could modify the behaviors of clinicians and increase the number of case reviews. Therefore we have added a sensitivity analysis to address the reviewer's concern. We will track when each medical center is notified that they failed to meet the targeted case review rate and we will have 3 groups: control, treatment (exposure to policy language) with facilitation, and treatment without facilitation. This analysis will allow us to estimate the effect of treatment with facilitation compared to control and treatment without facilitation. In addition, both our primary analyses and our sensitivity analyses include lagged treatment indicators to reflect diminishing returns of additional months of exposure to treatment.

- These changes have been made on p. 7 of the manuscript.

4. In the recruitment section on page 6, it would be nice to see an estimate of numbers of Veterans in the overall analytic cohort—in the top 10% of risk scores. See lines 22-23.

- We have added the estimated number of Veterans in the overall analytic cohort on page 6, line 22-23.
"Our analytic cohort will include approximately 100,000 VHA patients with an opioid prescription in the top 10% of risk scores."

Reviewer: 2
Reviewer Name: Michael Saenger

Institution and Country: Atlanta VA Health Care System, USA; , Emory University, USA

Bravo for asking and trying to answer these two questions, and for helping to raise the bar in care and in policy implementation.

- Thank you!

Reviewer: 3
Reviewer Name: André Scherag

Institution and Country: Institute of Medical Statistics, Computer Sciences and Documentation, Jena University Hospital, Jena, Germany

I was asked to focus on the methods/statistics of this proposal.
It is positive that the authors plan a randomized comparison to answer their two main research questions. Question 1 is relatively clear and the design matches to the research question. However, research question 2 and the operationalization differ - if the change of the included top is the aim, then this should be mirrored in the research question. Again, this also has an impact on the design and has to be reflected in both the sample size estimation (which should focus on the unit of randomization which are the VHA medical centers) and the analysis (which has to deal with both the center effects and the temporal pattern). What if interactions (between the experimental groups) exist? Moreover, the authors should comment on why they suggest using SAE rates (I guess some kind of cumulative incidence?) in the sample size estimation part and time-to-event analyses in the methods? That will likely confuse readers. Finally, I would also suggest reading CONSORT extensions for cluster randomized studies in order to be able to address their special reporting requirements.

- Research question 2 refers to the effectiveness of STORM, which we evaluate among patients with risk scores between 1% and 5%. At month 0, patients with risk scores between 1% and 5% do not appear in the STORM dashboard (control group). Nine months after randomization, patients in the 1%-5% risk stratum are displayed in the STORM dashboard at half of the medical centers (treatment group). Fifteen months after randomization, patients in the 1%-5% risk stratum are displayed at all medical centers.

- We clarify that our sample size calculation does reflect the medical center as unit of randomization as well as the two different interventions. In addition, we have modified the manuscript to ensure that the research questions and their operationalization match. We now include equations to represent our planned analyses on page 7. In our analysis of STORM effectiveness (Equation 2), we can test for an interaction between the policy indicator and the risk targeting indicator.

- Opioid-related SAEs are our primary outcomes of interest. Our sample size calculations reflect expected number of SAEs per person in a given month (not cumulative incidence). Thanks to the reviewer's comment, we double-checked our calculations and identified and corrected an error in our power calculation. However, we remain powered to detect clinically meaningful changes in SAE rates. We corrected the sample size/power calculation section on page 6 and modified the text in accordance with the CONSORT requirements for reporting cluster randomized study design.

VERSION 2 – REVIEW

REVIEWER	Michael Saenger, MD Atlanta VA Heath Care System; and Emory University
REVIEW RETURNED	16-Feb-2018

GENERAL COMMENTS	please clarify in table 1 that the "numerator" is not any STORM review in last 4 quarters but a review from the "denominator" group, i.e. "very high risk" during 7 day period in last 4 quarters

REVIEWER	Steven Dobscha MD VA Portland Health Care System Oregon Health & Science University Portland, Oregon, USA
REVIEW RETURNED	21-Feb-2018

GENERAL COMMENTS	I previously reviewed this manuscript, and the authors have address all of my comments and concerns (as well as those of other reviewers). My only remaining recommendation is with regard to Table 1: Here I would not recommend using the term Policy Memo Language, since we don't know what the specific language will be. Instead I might use the term Policy Memo Content which does not imply specific language. This is a very minor concern.

REVIEWER	André Scherag Institute of Medical Statistics, Computer and Data Sciences; Jena University Hospital; Germany
REVIEW RETURNED	22-Feb-2018

GENERAL COMMENTS	I thank the authors for addressing my concerns. However, two wishes and a comment remain. First of all, the titel of the manuscript should include that the study uses a stepped-wedge design. Second, the authors have added the SPIRIT 2013 checklist (Thanks!). However, some of the items are not adressed (for various reasons) and readers would like to see arguments for this deviation. I recommend adding these arguments to the supplement. Finally, I would like to mention that I did not double-check your sample size calculations (basically due to time constraints) but assume (given that you refere to a specific package) that it shoud be reproducible.

VERSION 2 – AUTHOR RESPONSE

In the manuscript, we have changed “Policy Memo” to “Policy Notice” to be consistent with a recent change in VA terminology. The content and the role of the Policy is exactly the same.

Reviewer(s)' Comments to Author:

Reviewer: 2
Reviewer Name: Michael Saenger, MD

Institution and Country: Atlanta VA Heath Care System; and Emory University

Please state any competing interests or state ‘None declared’: No financial competing interests

Please leave your comments for the authors below please clarify in table 1 that the "numerator" is not any STORM review in last 4 quarters but a review from the "denominator" group, i.e. "very high risk" during 7 day period in last 4 quarters.

Dr. Saenger,

Thank you very much for taking the time to review our manuscript thoroughly. You are correct to point out that the numerator is derived from the denominator cohort. To clarify, we have modified Table 1 to the following:

Denominator:

Patients with an opioid prescription who are in the “Very High – Opioid Patients” risk category in STORM for at least 7 days in the last quarter.

Numerator:

Patients in the denominator with case review within the last 4 quarters.

Reviewer: 1
Reviewer Name: Steven Dobscha MD

Institution and Country: VA Portland Health Care System, Oregon Health & Science University, Portland, Oregon, USA

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below I previously reviewed this manuscript, and the authors have address all of my comments and concerns (as well as those of other reviewers). My only remaining recommendation is with regard to Table 1: Here I would not recommend using the term Policy Memo Language, since we don't know what the specific language will be. Instead I might use the term Policy Memo Content which does not imply specific language. This is a very minor concern.

Dr. Dobscha,

Thank you very much for taking the time to review our manuscript thoroughly. We have modified the Table 1 to say Policy Notice Content in page 4.

Reviewer: 3
Reviewer Name: André Scherag

Institution and Country: Institute of Medical Statistics, Computer and Data Sciences; Jena University Hospital; Germany

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below I thank the authors for addressing my concerns. However, two wishes and a comment remain. First of all, the title of the manuscript should include that the study uses a stepped-wedge design. Second, the authors have added the SPIRIT 2013 checklist (Thanks!). However, some of the items are not addressed (for various reasons) and readers would like to see arguments for this deviation. I recommend adding these arguments to the supplement. Finally, I would like to mention that I did not double-check your sample size calculations (basically due to time constraints) but assume (given that you refer to a specific package) that it should be reproducible.

Dr. Scherag,

Thank you very much for taking the time to review our manuscript thoroughly. We have modified our title to:

The Effectiveness of Policy and Risk Targeting for Opioid-Related Risk Mitigation: A Randomized Program Evaluation with Stepped-Wedge Design

To address your 2nd comment, we have added some explanation to each checklist item. Since this protocol is not for a clinical trial, some of the questions were not relevant and were marked as "NA".

We believe the sample size calculation is reproducible using the appropriate State packages.