PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Reducing the default dispense quantity for new opioid analgesic prescriptions: study protocol for a cluster randomized controlled trial
AUTHORS	Bachhuber, Marcus; Nash, Denis; Southern, William; Heo, Moonseong; Berger, Matthew; Schepis, Mark; Cunningham, CO

VERSION 1 – REVIEW

REVIEWER	Michael Zwank Regions Hospital Saint Paul, MN USA University of Minnesota Minneapolis, MN USA
REVIEW RETURNED	25-Sep-2017

GENERAL COMMENTS	This is a well-written description of a research study protocol. The study is well-conceived, well-designed and seems likely to answer the authors' question as proposed. I don't have any specific edits or general concerns. I will be excited to see the results of the study as a reviewer or as a interested clinician reading the final publication. Some of the proposed statistical analysis is beyond my scope and I would defer to a trained statistician to comment on this.

REVIEWER	Christopher Harle Indiana University, Indianapolis (IUPUI), USA
REVIEW RETURNED	02-Oct-2017

GENERAL COMMENTS	General Comments: This manuscript is a well-written Protocol article describing a forthcoming trial to examine the use of default opioid prescribing quantities defined in electronic health records (EHRs). Given the public health crisis surrounding opioid misuse, abuse, use disorder, and diversion in the United States, this is an important study. Moreover, given the near ubiquitous use of EHRs in the United States, EHRs represent a sensible and important medium for delivering interventions meant to change clinician decision making. The study design is strong and appropriate. The study is limited by observing only a single health system, which weakens generalizability and comes with an inability to examine outside system utilization as well as longer term outcomes. However, the article recognizes these limitations. Overall, the article would benefit from clearer descriptions and justifications in several areas as described below. Specific Comments: 1. Many states have recently implemented laws that cap opioid

	prescription quantities. The article should comment on this, including any NY State laws and how such restrictions could affect the planned study. 2. The study plans to exclude patients who have received an opioid prescription in the preceding 6 months (p6 line 25). It would help to clarify if that exclusion applies to all opioids, including extended release and/or formulations that are not listed among those to which the default quantity is being applied (p6 lines 42). For example, will patients with chronic pain receiving an ER/LA opioid potentially be included? 3. Will individual patients be able to appear in the data multiple times (i.e., if the receive a second new prescription in the study period)? Please clarify/justify. 4. Please describe if any primary care or ED sites changed their defaults in the recent time leading up to the study. 5. The analysis might benefit from also collecting history of chronic pain conditions and including as covariates along with the other planned history covariates. While the study focuses on opioid prescriptions for acute pain, it seems that patients with chronic pain conditions will be present in the sample, in particular those not receiving chronic opioid therapy. 6. Please explain in more detail how pain diagnosis and illness histories will be measured (p8, line 40). 7. A sentence or two better justifying the chosen strata for the randomization would be helpful. 8. Some alternate modeling of the DID could help further explore the effects of defaults, including whether their effect are immediate and/or if and when their effects decay over time. The authors might consider additional secondary analyses to explore this. Similarly, the robustness of any observed effects could be examined with additional sensitivity analyses around their timing. 9. The analysis should address the parallel trends assumption underlying a DID analysis and how the strength of this assumption will be examined. 10. It would help to provide more practical explanation of the plan to explore heterogeneity of the intervention's effects between matched pairs and why it is important (p 10, line 53). 11. Please justify the sample size assumption of 3% increase in <=10 day prescriptions for the control group. One might speculate that the general trend toward decreased opioid prescribing could be even more rapid. Would this affect the sample size required? 12. The limitations would benefit from a more detailed description of the weaknesses inherent in cluster randomization and DID.

REVIEWER	Katherine Morley King's College London, United Kingdom
REVIEW RETURNED	01-Nov-2017

GENERAL COMMENTS	This is an interesting study, and the statistical approach the authors plan to use appears to be appropriate for the research question and the data collected. However, I found that the Methods section was lacking in detail in some areas, which I have outlined below. [1] On page 8, the authors specify a number of variables which may not necessarily be straightforward to extract from the EHR: (a) secondary outcome 2: health service utilisation within 30 days related to opioid use; (b) pain disorder (indication for prescription); (c) history of substance use disorder; (d) history of psychiatric diagnoses. Deriving this information can be complex, and may warrant study in its own right. Are the authors using pre-existing algorithms for extracting this information? If not, please provide more information about how how these concepts will be defined. Is this via the use of structured data (e.g. ICD-10 codes) or via natural language processing of clinical notes, or some other method? If codes are to be used, please provide further information (this could be included as an appendix/supplementary, rather than in the main text if the list is long). [2] On page 10 the authors state: “In addition to this specification, we will explore methods to allow for heterogeneity of the intervention’s effect between matched pairs.” Please provide more information about the methods that will be explored. [3] Testing the assumptions of the difference-in-differences analysis has not been explicitly addressed in the current protocol. A good discussion of these issues is provided by Ryan et al. (2015) Why we should not be indifferent to specification choices in Difference-in-Differences. HSR: Health Services Research 50(4):1211. I highlight the main points below: (a) One of the key assumptions of the DiD approach is that the comparison provides an appropriate counterfactual for the intervention group. To evaluate whether this assumption is likely to hold, an investigator needs data for at least two time-points prior to the intervention. In the current protocol, the authors do not mention doing this, but presumably given that they have access to EHR data this would be feasible? (b) Another assumption is the absence of “spill over” from treatment group to comparison group. On page 9, the authors state that “virtually all providers only practice at one site”, but for the purposes to quantifying the potential for spill over effects, it may be helpful for the authors to consider identifying how many providers do practice across sites and how this could affect the results. (c) For DiD analyses, it is also important to ensure that the composition of the intervention and comparison groups is stable over time - do the authors have plans for how to address this? (d) Ryan et al. suggest that clustered standard errors and permutation tests should be used in DiD analyses to handle violations of independence - do the authors plan to implement any of these approaches?

VERSION 1 – AUTHOR RESPONSE

1. Editorial Requirements:
- Please ensure the manuscript is correctly formatted as per our guidelines for protocol articles: http://bmjopen.bmj.com/pages/authors/ For example, please remove the conclusions section.
- Please clarify when participant enrollment began in the manuscript.

•We have now removed the Conclusions section, worked the Limitations into the "Methods and analysis" section and removed several bolded headings in the "Methods and analysis". In addition, we have clarified the date of the EHR change in the "Timeline and monitoring" subsection of the "Methods and analysis".

Reviewer 1
Reviewer Name: Michael Zwank

2. This is a well-written description of a research study protocol. The study is well-conceived, well-designed and seems likely to answer the authors' question as proposed. I don't have any specific edits or general concerns. I will be excited to see the results of the study as a reviewer or as a interested clinician reading the final publication. Some of the proposed statistical analysis is beyond my scope and I would defer to a trained statistician to comment on this.
•We appreciate the reviewer's kind words. The other two reviewers have suggested some further modification and specification of the statistical analysis sections, which we have addressed below.

Reviewer 2
Reviewer Name: Christopher Harle
General Comments:
This manuscript is a well-written Protocol article describing a forthcoming trial to examine the use of default opioid prescribing quantities defined in electronic health records (EHRs). Given the public health crisis surrounding opioid misuse, abuse, use disorder, and diversion in the United States, this is an important study. Moreover, given the near ubiquitous use of EHRs in the United States, EHRs represent a sensible and important medium for delivering interventions meant to change clinician decision making. The study design is strong and appropriate. The study is limited by observing only a single health system, which weakens generalizability and comes with an inability to examine outside system utilization as well as longer term outcomes. However, the article recognizes these limitations. Overall, the article would benefit from clearer descriptions and justifications in several areas as described below.

•We thank the reviewer for the kind words and are extremely grateful for the time the reviewer invested in providing specific and constructive feedback.

Specific comments:
3. Many states have recently implemented laws that cap opioid prescription quantities. The article should comment on this, including any NY State laws and how such restrictions could affect the planned study.

•We agree with the reviewer that this is a highly relevant policy and have now included information about these laws in the Introduction (paragraph 2, last sentence): "In addition, as of December 2017, 24 states have passed laws setting limits on new opioid analgesic prescriptions; however, enforcement mechanisms are often unclear and the impact of such laws on prescribing is not known."

•In addition, we have added details about New York State's acute pain/opioid prescribing law in the Statistical analysis section (paragraph 2, last sentence) when discussing our choice of DID: "For

example, in July 2016, New York State enacted a law limiting opioid analgesic prescriptions for acute pain to a 7-day supply."

4. The study plans to exclude patients who have received an opioid prescription in the preceding 6 months (p6 line 25). It would help to clarify if that exclusion applies to all opioids, including extended release and/or formulations that are not listed among those to which the default quantity is being applied (p6 lines 42). For example, will patients with chronic pain receiving an ER/LA opioid potentially be included?

•The reviewer raises an excellent point. We plan to exclude those receiving any opioid prescription in the preceding 6 months. We have now clarified this under "Eligibility criteria", subsection "Patient participants" (new text bolded): "We will analyze outcomes for patients that: a) received a new opioid analgesic prescription, defined as no other opioid analgesic prescription of any type in the preceding 6 months (a definition used in previous cohort studies)…"

5. Will individual patients be able to appear in the data multiple times (i.e., if the receive a second new prescription in the study period)? Please clarify/justify.

•We agree that this needs clarification and have now added a new sentence to the "Eligibility criteria", subsection "Patient participants": "For patients receiving more than one new opioid analgesic prescription during the study period, we will only include the first prescription." We are making this limitation because we are interested in new opioid analgesics for acute pain, not repeat prescriptions.

6. Please describe if any primary care or ED sites changed their defaults in the recent time leading up to the study.

•We have now added two sentences describing this in more detail ("Timeline and monitoring" subsection, first paragraph), "Before this change, primary care sites had the same EHR for approximately 19 months. Two emergency department sites had the same EHR for 11 months, and two emergency department sites had the same EHR system for 7 months (i.e., those sites implemented the current EHR just before start of the 6-month pre-intervention period)."

7. The analysis might benefit from also collecting history of chronic pain conditions and including as covariates along with the other planned history covariates. While the study focuses on opioid prescriptions for acute pain, it seems that patients with chronic pain conditions will be present in the sample, in particular those not receiving chronic opioid therapy.

•We agree completely with the reviewer's point that chronic pain conditions may be important confounders; however, we are concerned at our ability to reliably distinguish acute versus chronic pain from the medical record. We currently do not have the ability to do natural-language processing of medical records, so we will identify conditions using ICD-10-CM diagnosis codes. Based on the available codes and our experience of coding practices at our institution, we do not believe we can reliably determine which patients have acute on chronic pain or simply acute pain.

8. Please explain in more detail how pain diagnosis and illness histories will be measured (p8, line 40).

•We agree that more specificity is necessary and have now updated the subsection "Provider and patient characteristics (covariates)". For pain conditions, we will adapt the clinical categories outlined in the National Pain Strategy and map ICD-10-CM diagnosis codes into these categories. Our groupings of codes will be based on published literature (which is quite limited now, but increasing) and actual coding patterns at our institution (i.e., a data-driven approach). We realize that pre-

specifying codes at the time of publication of the study protocol would be ideal but, having reviewed the literature, we do not believe there is enough information currently available. We also believe that grouping codes into clinically-meaningful categories may depend, at least in part, on the codes actually used at our institution (which we can only evaluate retrospectively). Therefore, as this is a covariate and not a main outcome, we have opted to provide a structured framework for how we will define this variable instead of pre-specifying the exact codes we will use.

•For history of psychiatric illness and history of substance use disorder, we will use existing diagnosis code groupings produced by the Healthcare Cost and Utilization Project, sponsored by the Agency for Healthcare Research and Quality of the United States Department of Health and Human Services. We have edited the text of the paragraph to describe this and have included the reference.

9. A sentence or two better justifying the chosen strata for the randomization would be helpful.

•We agree and have now edited the second paragraph of the "Randomization" subsection (starting second sentence) to read, "We will stratify sites by type (i.e., primary care versus emergency department). Further, within primary care sites, prescribing patterns and the intervention's impact may differ by specialty (i.e., internal medicine and family medicine) and whether the site is a training site for resident physicians. Therefore, we will stratify on these variables as well."

10. Some alternate modeling of the DID could help further explore the effects of defaults, including whether their effects are immediate and/or if and when their effects decay over time. The authors might consider additional secondary analyses to explore this. Similarly, the robustness of any observed effects could be examined with additional sensitivity analyses around their timing.

•For the first point, as part of our analysis plan we are proposing to analyze outcomes from -6 to +6 months as well as outcomes at 18 months. We have highlighted the issue of potential decay further (Statistical analysis section, fifth paragraph, last sentence, new text bolded), "When analyzing the impact of the intervention at 18 months, we will identify any change in the intervention's impact after 6 months (i.e., whether it decays over time) by using the 0 to 6 month post-intervention period as the referent."'

•On the second point, although we agree completely that additional secondary and sensitivity analyses would provide further information, we have already described a large number of analyses in this protocol (several outcomes, some with more than one measure, at two time points, with pre-specified subgroup analyses). In addition, due to resource constraints, we are hesitant to commit to analyses that we may not be able to carry out. We have now edited the text to reflect this and created a separate paragraph ("Statistical analysis" section, seventh paragraph, new text bolded), "Finally, we intend to explore other analyses examining the precise timing of any changes in outcomes (e.g., immediate or delayed) and to characterize the heterogeneity of the intervention's effect between matched pairs. Such analyses will be defined post-hoc and are subject to availability of resources such as additional statistical support and technical considerations such as convergence of relevant statistical models."

11. The analysis should address the parallel trends assumption underlying a DID analysis and how the strength of this assumption will be examined.

•We agree and have added detail about how we will address the parallel trends assumption. We have worked this in to a larger paragraph about DID assumptions (Statistical analysis section, third paragraph, see Reviewer 3, #17).

12. It would help to provide more practical explanation of the plan to explore heterogeneity of the intervention's effects between matched pairs and why it is important (p 10, line 53).

•We agree that, as written, it appeared incomplete. We have now changed how we characterize these analyses, given that they will be developed post-hoc and are subject to availability of resources. Our changes are detailed in #10 above.

13. Please justify the sample size assumption of 3% increase in <=10 day prescriptions for the control group. One might speculate that the general trend toward decreased opioid prescribing could be even more rapid. Would this affect the sample size required?

•The reviewer raises an excellent point that this assumption was arbitrary. There are no data to guide what we may expect in the control arm. Therefore, we have now provided estimates of the minimal detectable difference for a range of increases in the control arm from 0 to 10% (new text bolded): "Because any change in outcomes in the control arm is also unknown, we used a range of increases in the percentage of prescriptions for ≤ 10 tablets in the control arm of between 0 and 10 percentage points. Within this range of ICC, change in the control arm, alpha=0.05, and power ≥ 80%, this study will be powered to detect a change in the intervention arm, over and above any change in the control arm, of 4.4 to 4.7%."

•We considered putting these calculations in a table, but given that the range is only 4.4 to 4.7%, the table appeared very repetitive and did not add to the text description.

14. The limitations would benefit from a more detailed description of the weaknesses inherent in cluster randomization and DID.

•We agree that expanding our discussion of the design considerations is necessary. For cluster randomization, we have now included the major trade-off of this design ("Randomization" subsection, new text bold): "Compared to randomization at the level of the provider (i.e., individual-level randomization), randomization of sites would be expected to reduce statistical efficiency due to correlated outcomes within clusters.(Donner 2004) However, we chose site-level randomization instead of provider-level randomization to reduce contamination and to potentially increase the intervention's effectiveness via peer effects."

•For DID, we have now included a paragraph addressing its assumptions and how we will test them ("Statistical analysis" subsection, third paragraph, see Reviewer 3, #17).

Reviewer: 3
Reviewer Name: Katherine Morley
This is an interesting study, and the statistical approach the authors plan to use appears to be appropriate for the research question and the data collected. However, I found that the Methods section was lacking in detail in some areas, which I have outlined below.

•We are extremely appreciative of the time the reviewer invested into identifying specific weaknesses that we can address to strengthen the manuscript.

15. On page 8, the authors specify a number of variables which may not necessarily be straightforward to extract from the EHR:
(a) secondary outcome 2: health service utilisation within 30 days related to opioid use;
(b) pain disorder (indication for prescription);
(c) history of substance use disorder;
(d) history of psychiatric diagnoses.

Deriving this information can be complex, and may warrant study in its own right. Are the authors using pre-existing algorithms for extracting this information? If not, please provide more information about how how these concepts will be defined. Is this via the use of structured data (e.g. ICD-10 codes) or via natural language processing of clinical notes, or some other method? If codes are to be used, please provide further information (this could be included as an appendix/supplementary, rather than in the main text if the list is long).

•We agree and have further clarified how we will obtain these data. For (a), we will not be attempting to classify whether health service utilization will be opioid-related. We have clarified this under "Secondary outcomes" (new text bolded): "We will analyze the number of primary care visits, ED visits, and hospitalizations for any reason."

For (b) through (d), we have further specified how we will define these variables above (Reviewer 2, #8).

16. On page 10 the authors state: "In addition to this specification, we will explore methods to allow for heterogeneity of the intervention's effect between matched pairs." Please provide more information about the methods that will be explored.

•We agree that our description was incomplete, and we have reframed this statement, described under Reviewer 2, #10 and #12 above.

17. Testing the assumptions of the difference-in-differences analysis has not been explicitly addressed in the current protocol. A good discussion of these issues is provided by Ryan et al. (2015) Why we should not be indifferent to specification choices in Difference-in-Differences. HSR: Health Services Research 50(4):1211. I highlight the main points below:
(a) One of the key assumptions of the DiD approach is that the comparison provides an appropriate counterfactual for the intervention group. To evaluate whether this assumption is likely to hold, an investigator needs data for at least two time-points prior to the intervention. In the current protocol, the authors do not mention doing this, but presumably given that they have access to EHR data this would be feasible?
(b) Another assumption is the absence of "spill over" from treatment group to comparison group. On page 9, the authors state that "virtually all providers only practice at one site", but for the purposes to quantifying the potential for spill over effects, it may be helpful for the authors to consider identifying how many providers do practice across sites and how this could affect the results.
(c) For DiD analyses, it is also important to ensure that the composition of the intervention and comparison groups is stable over time - do the authors have plans for how to address this?
(d) Ryan et al. suggest that clustered standard errors and permutation tests should be used in DiD analyses to handle violations of independence - do the authors plan to implement any of these approaches?

•The reviewer raises excellent points and we have now greatly strengthened the description of how we plan to test the assumptions of DID. We have added a new paragraph ("Statistical analysis" subsection, third paragraph): "A DID analysis also relies on several assumptions which we will examine.(Ryan 2015 reference) First, we will assess whether trends in outcomes were parallel between the intervention and control sites prior to the intervention. For this analysis, in the pre-intervention period, we will determine the significance of an interaction term between study allocation (intervention/control) and month. Second, to determine the composition of the intervention and the control sites, we will calculate and report descriptive statistics for both provider and patient characteristics, pre- and post-intervention. Finally, we will examine the potential for contamination of

the arms. Although we expect the number of providers that write prescriptions at both an intervention and a control site will be low, we will determine the number of such providers and report it."

•For the composition issue, we have slightly modified the language in the paragraph about the main DID analysis (Statistical analysis" subsection, fourth paragraph, third sentence, new text bolded): "To adjust for potential changes in composition over time, we will include relevant site characteristics (number of new opioid analgesic prescriptions, the number of visits, and percentage of patients with commercial insurance), provider characteristics (sex and years since medical school graduation) and patient characteristics (age, sex, race/ethnicity, pain diagnosis, history of substance use disorder, history of psychiatric disorder) as covariates in all models."

•Finally, for (d), non-independence is being accounted for by using random intercepts at both the provider and patient levels. In addition, we will calculate robust standard errors and we have added this ("Statistical analysis" section, fourth paragraph, last sentence): "For all estimates, we will calculate heteroscedasticity robust (empirical) standard errors."

VERSION 2 – REVIEW

REVIEWER	Christopher Harle Indiana University Richard M. Fairbanks School of Public Health, USA
REVIEW RETURNED	17-Feb-2018

GENERAL COMMENTS	The authors have done an excellent job of responding to the first round of reviewer comments, making many helpful edits to clarify their study design and analysis. I look forward to seeing the study results.