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)	Measuring the relationship between interruptions, multi-tasking and prescribing errors in an emergency department: study protocol
AUTHORS	Raban, Magdalena; Walter, Scott; Douglas, Heather; Strumpman, Dana; Mackenzie, John; Wesbrook, Johanna

VERSION 1 - REVIEW

REVIEWER	Matthias Weigl / Anna Schneider Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine Munich Unversity Hospital Ludwig-Maximilians-University Munich Germany
REVIEW RETURNED	14-Jul-2015

GENERAL COMMENTS	The paper presents a protocol of a study, designed to investigate the associations of ED clinicians' workflow interruptions, multitasking activities and prescribing errors. Since particularly ED work is prone to a high number of interruptions and break-in tasks, this setting is very well suitable to investigate the impact of deficient work practices, i.e., interruptions and multitasking demands, and potential effects for unsafe clinical care, i.e., prescribing errors. Beyond its relevance, the particular strengths of the study refer to the use of multi-source data, the application of well-established observational tools, as well as its attempt to identify prescription errors based on experts' evaluations. Overall this is a well drafted study protocol with a well-described background and well drafted descriptions of the respective study steps. Since the overall quality of the manuscript is very good, we have only one major issue that should be addressed to a farther extent in a revised version and a few minor issues: 1) Inclusion of a paragraph of potential limitations of the study. Currently just a few notes are devoted to potential limitations of the study. However, we deem that potential limitations should be discussed beforehand. We assume that particularly two limitations might limit the validity of the study. First, we understood that this is a cross-sectional study. This limits inferences about causality of the interruptions/multitasking – errors relationship. Secondly, we wondered how you will deal with observer bias. Particularly if observed clinicians are aware about the study objectives. Could it be that observed physicians (and co-workers) behave less distracting or pay more attention to prescriptions since they know the study

	objectives? How may this interfere with your study objectives? Discussion on further limitations that may limit the external validity may include the focus on solely one ED in Australia and the focus on day shifts. Minor issues: - Page 6, line 30f. Readers might not be familiar with the Australian ED care system. Could you please shortly describe which qualification or specialty the observed physicians will most likely have? - Additionally, it might be relevant for foreign readers to what extent medications are prescribed in Australian EDs. Particularly, if there are differences between ED patients who are admitted to the hospital (e.g., ICUs, ORs, inpatient wards) and those who are sent back home (or to a GP) after their ED visit. - Page 6, lines 54ff. How will you select the observation dates? Randomly? Will you include also weekends and/or holidays? - Page 6, lines 54ff. We also wondered if you considered examining night shifts as well. We assume that the exclusive focus on day shifts was based on specific reasons. Could you please expand on that? - Page 7, lines 3ff. What training do observers receive prior to the study; how do you ensure that observer agreement is established? - Page 10, line 51. What do you mean by "… and a binary variable of interest was assumed to occur 50% of the time?" This is not clear yet. - Page 11, line 31, "…per unit time". It is not clear which unit time is considered here. - Page 11, line 35. We wondered if it could be necessary to adjust also for the number of patients seen that shift. Since crowding or overcrowding has been shown to severely impact the quality of ED care, this number could serve as a proxy for work overload or time pressure. We hope that the issues mentioned above will help to improve the current draft and study protocol. We wish the authors good luck with their manuscript. We look forward to read about the results of the study.

REVIEWER	Nina Dadlez, MD The Children's Hospital at Montefiore, USA
REVIEW RETURNED	29-Jul-2015

GENERAL COMMENTS	This is an interesting and relevant protocol studying interruptions in the emergency department setting and their relationship to medication prescribing errors. It is a well-written and well-designed protocol. I have several comments for the authors: 1. Please include a discussion of the Hawthorne effect in your limitations section. As it would be impossible for structured observations to occur without the study participant's knowledge that they are being observed, this may directly influence his/her behavior. 2. Please acknowledge that this is an emergency department that does not have computerized physician order entry (CPOE) in the limitations section. This would limit generalizability to institutions that do have CPOE. Several of the definitions of medication prescribing

	error listed in appendix 1, particularly in the procedural and legal sections, would not be applicable in a CPOE environment. 3. I would consider calculating the prescribing error rate utilizing number of orders placed as the denominator. While prescribing error rates are reported in variable manners in the literature (percentages, errors per 100 orders, errors per 100,000 orders), the denominator is typically the number of orders placed. Utilizing prescribing task may be harder for readers to compare with existing literature. If you choose to utilize prescribing task, I recommend including a sentence or two explaining this reasoning to the reader. 4. In the statistical analysis section on page 10 lines 48-50, I would recommend clarifying that you are assessing effect sizes for detection of error rates. 5. In appendix 1 under "clinical errors: not prescribed," please provide an example of how you will determine if a medication that is clinically indicated is not prescribed. 6. Page 4 line 48 is a sentence fragment. Please consider writing "errors are" instead of "error is". 7. Page 8 line 39-42: Is reference #37 a peer-reviewed publication or an unpublished protocol? If the latter, please disclose that within the manuscript. In addition, reference #38 does not mention the error classification schematic. References are otherwise up to date and appropriate.

VERSION 1 – AUTHOR RESPONSE

Reviewers' comments	Responses to comments
Reviewer #1	
Inclusion of a paragraph of potential limitations of the study. Currently just a few notes are devoted to potential limitations of the study. However, we deem that potential limitations should be discussed beforehand. We assume that particularly two limitations might limit the validity of the study. First, we understood that this is a cross-sectional study. This limits inferences about causality of the interruptions/multitasking – errors relationship. Secondly, we wondered how you will deal with observer bias. Particularly if observed clinicians are aware about the study objectives. Could it be that observed physicians (and co-workers) behave less distracting or pay more attention to prescriptions since they know the study objectives? How may this interfere with your study objectives? Discussion on further limitations that may limit the external validity may include the focus on solely one ED in Australia and the focus on day shifts.	A 'Limitations' section has been added, before the 'Ethics and Dissemination', section on pages 12-13 covering the points raised by the reviewer. The cross-sectional nature of the study design is debatable as observations session will take place over the course of several months and participants will be observed for multiple sessions. Also it is not clear what is meant by a cross-sectional design being a limitation for causality. Establishing causality between interruptions/multitasking and errors is incredibly challenging and to date no one has managed to do this. Our study aims to assess association at a fine grained level (within prescribing tasks) which helps to establish the temporal proximity of interruptions/multitasking and error (a condition of causality). Furthermore, we will attempt to account for other potentially confounding factors in the analysis which addresses another condition of causality – eliminating alternative explanations. We appreciate that there is scope to do more to try

	to establish a causal link, however, this study design makes inroads in this direction beyond what previous studies have managed. We have extensive experience in applying this observational approach and have found that significant, prolonged changes in clinicians' behaviour in busy clinical environment is unlikely. There is also some evidence from other studies that clinical staff do not significantly change their behaviours when observed. (See for example: Dean B, Barber N. Validity and reliability of observational methods for studying medication administration errors. Am J Health Syst Pharm 2001;58(1):54-9.)
Page 6, line 30f. Readers might not be familiar with the Australian ED care system. Could you please shortly describe which qualification or specialty the observed physicians will most likely have?	We have added text to this section in order to clarify this point. This section now reads (p. 6): "A qualified doctor in Australia has completed a medical degree and one year of on the job training. This includes resident medical officers (RMO; 1-2 post-graduate years' experience), senior resident medical officers (SRMO; 2-3 post-graduate years' experience), registrars (at least 3 post-graduate years' experience and completion of relevant training) and staff specialists (at least 5 post-graduate years' experience and completion of relevant training)."
Additionally, it might be relevant for foreign readers to what extent medications are prescribed in Australian EDs. Particularly, if there are differences between ED patients who are admitted to the hospital (e.g., ICUs, ORs, inpatient wards) and those who are sent back home (or to a GP) after their ED visit.	Prescribing practices in Australian EDs are explained in a paragraph added at the end of the "Study settings and participants" section (p. 6-7): "All qualified doctors listed above can prescribe medications. In Australian EDs, doctors prescribe medications that need to be administered in the ED, as well as on wards when a patient is admitted to the hospital. In the acute section of the study ED, approximately 60-70% of patients are admitted."
Page 6, lines 54ff. How will you select the observation dates? Randomly? Will you include also weekends and/or holidays?	It is essentially impossible to apply a planned (randomised or not) timetable of observation sessions. Doctors may not give consent at a particular time, a doctor who has given consent may swap shift at the last minute, and so on. Furthermore, attempting to capture day shift, evening shift and night shift across weekdays and weekends is a formidable task requiring a vast increase in sample size and study

	resources. For these logistical reasons we will limit observations to weekday day shifts. Observation days will initially be selected according to observer availability and as observations progress sessions will be chosen to achieve approximate balance across doctor groups and time of day.
Page 6, lines 54ff. We also wondered if you considered examining night shifts as well. We assume that the exclusive focus on day shifts was based on specific reasons. Could you please expand on that?	We agree with the reviewer that including night shifts would be an interesting addition, especially since working on night shift adds another stressor to doctors' work. However, as outlined above, this would require considerable resources. Thus, due to resource constraints for this study, we have limited our sample to day shift.
Page 7, lines 3ff. What training do observers receive prior to the study; how do you ensure that observer agreement is established?	Observers will undergo extensive training with the data collection tool and data collection will not commence until inter-rater reliability (IRR) testing results show a high level of agreement. Additionally, IRR will be conducted at several time points during data collection to ensure agreement is maintained. Text has been added to page 8: "Observers will undergo extensive training in using the data collection tool. Prior to starting data collection, inter-rater reliability (IRR) testing will be undertaken and data collection will only begin once observers have reached high agreement. IRR will be reassessed at several time points during the data collection period to ensure consistency over time between observers."
Page 10, line 51. What do you mean by "… and a binary variable of interest was assumed to occur 50% of the time?" This is not clear yet.	We have revised this to clarify the intended meaning as follows: "Calculations were performed for multivariate Poisson regression with a main binary independent variable of interest (e.g. was the prescribing task interrupted or not), assumed to occur in 50% of prescribing tasks, with the R^2 of the other covariates assumed to be in the range 0.1 to 0.5. This is a slightly simplified version of the Poisson model, compared to the proposed analysis, in order to facilitate the sample size calculation." (p. 11)
Page 11, line 31, "…per unit time". It is not clear	"Per unit time" is a standard expression indicating that time is the denominator of a rate.

which unit time is considered here.	In terms of including rates of task-switching and multitasking per unit time in the Poisson regression model, any particular unit can be used (minutes, hours, etc.) as long as the interpretation of model estimates is consistent with the chosen unit.
Page 11, line 35. We wondered if it could be necessary to adjust also for the number of patients seen that shift. Since crowding or overcrowding has been shown to severely impact the quality of ED care, this number could serve as a proxy for work overload or time pressure.	This is a good suggestion and it is possible to access the EDs data on patient load. There are many ways to measure individual and departmental ED workload. A more fine-grained department level measure is the number of patients in the ED at the time a prescribing task is observed. We have included this as a covariate in the analysis as follows: "We will also adjust for other factors including patient age, doctor seniority and patient load at the time of the prescription. Patient load data will be extracted from the ED information management system." (p.12)
Reviewer #2	
Please include a discussion of the Hawthorne effect in your limitations section. As it would be impossible for structured observations to occur without the study participant's knowledge that they are being observed, this may directly influence his/her behavior.	We have added a section titled 'Limitations' which discusses the Hawthorne effect. (p.12-13) Please also refer to response to Reviewer 1's first comment for more detail.
Please acknowledge that this is an emergency department that does not have computerized physician order entry (CPOE) in the limitations section. This would limit generalizability to institutions that do have CPOE. Several of the definitions of medication prescribing error listed in appendix 1, particularly in the procedural and legal sections, would not be applicable in a CPOE environment.	We have added this acknowledgement in the 'Limitations' section. It is also important to note that our analysis will use prescribing error rates based on all errors, and clinical, legal and procedural errors separately (p.13).
I would consider calculating the prescribing error rate utilizing number of orders placed as the denominator. While prescribing error rates are reported in variable manners in the literature (percentages, errors per 100 orders, errors per 100,000 orders), the denominator is typically the number of orders placed. Utilizing prescribing task may be harder for readers to compare with existing literature. If you choose to utilize prescribing task, I recommend including a sentence or two explaining this reasoning to the reader.	The error rate is indeed calculated as the number of errors divided by the number of orders. This is necessarily done at the task-level for the Poisson regression approach. We have modified the second and fourth sentences in this paragraph to clarify these details (p.11): "The prescribing error rate for a given prescribing task is the number of errors divided by the number of medication orders."

	Also (p.12): "Multivariate Poisson regression will be applied to the number of errors per task with the number of orders per task as the offset (i.e. denominator), hence modelling the error rate at task level."
In the statistical analysis section on page 10 lines 48-50, I would recommend clarifying that you are assessing effect sizes for detection of error rates.	This sentence has been modified: "Detectable effect sizes for differences in error rates were assessed for a sample size of 360 …" (p.11)
In appendix 1 under "clinical errors: not prescribed," please provide an example of how you will determine if a medication that is clinically indicated is not prescribed.	An example would be for a patient admitted with chest pain, where hospital protocol indicates they should receive a single dose of aspirin 300mg as soon as possible. If this did not occur, this would be a 'not prescribed' error. However, the prescribing error definitions presented in the appendix were originally designed for a prospective study of prescribing errors on wards which would examine all medication orders for a patient. In our context, the assessment for errors will only pertain to orders written at the time of observation, which is not likely to be all orders. Thus, the 'not prescribed' error is likely to be irrelevant, since we will not be able to assess whether a medication should have been prescribed at that particular point in time. Therefore, we have removed this error category.
Page 4 line 48 is a sentence fragment. Please consider writing "errors are" instead of "error is".	The sentence has been edited as suggested.
Page 8 line 39-42: Is reference #37 a peer-reviewed publication or an unpublished protocol? If the latter, please disclose that within the manuscript. In addition, reference #38 does not mention the error classification schematic. References are otherwise up to date and appropriate.	The text "Unpublished Protocol" has been added to reference number 37. Reference 38 has been removed.

VERSION 2 – REVIEW

REVIEWER	Anna Schneider / Matthias Weigl Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine Ludwig-Maximilians University Munich Munich Germany
REVIEW RETURNED	03-Sep-2015

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.

REVIEWER	Nina Dadlez, MD The Children's Hospital at Montefiore, USA
REVIEW RETURNED	02-Sep-2015

GENERAL COMMENTS	This is an interesting and relevant protocol studying interruptions in the emergency department setting and their relationship to medication prescribing errors. The time study design will be important to establish associations between interruptions and errors which has been difficult to do to date. It is a well-written and well-designed protocol. The authors have addressed the reviewer comments adequately and I appreciate the changes made. My only concern is the rapid dismissal of the Hawthorne effect in an observational study such as the protocol describes. The authors cite two references, however there many that report the contrary in clinical settings (Kidwai et al, Hagel et al, etc). I do think that the language included in the limitations section regarding this issue is adequate.