Author Reply to Reviewer 1

Reviewer 1:
This study tests the authors' theory that experienced physicians will adhere to standardized guidelines or algorithms less closely when specific patient circumstances suggest an exception might be in order. They also designed the study to compare those experienced physicians with less experienced physicians, hypothesizing that those less experienced will make fewer exceptions to the "rules" that help them reduce their diagnostic or therapeutic uncertainty. Their results largely confirmed their hypotheses, demonstrating a reduction in CPG adherence in the presence of these circumstances and a greater reduction in adherence among experienced vs. novice physicians.

Writing to the authors:

1) Introduction/background. This section generally sets up well the premise behind CPGs and why they might not be adhered to; however, the technical language does make for some dense reading. You could actually use just a few more sentences to explain why reduction in variation does generally improve quality and efficiency (replacing difficult last sentence of first paragraph) and offering an example of choice conditioning. The citations for the end of the 2nd paragraph seem to need to go at the end of the sentence ending CPG's, and again it would help to explain "patient preference" or distinguish that from "choice conditioning." It is challenging to choose a name for "extenuating circumstances" of the patient, but it is not obvious that's what "context variable" means in the abstract and that word isn't used that way until well into the methods. It would help the reader to define your term early in the paper as you offer examples in the introduction, and more specific examples could probably help the reader. The hypotheses are clear, though you might want to add to number three that you are assuming that both experienced and novice physicians are aware of the CPG and also testing to see if they generally adhere to it in principle.

Response: The introduction of the manuscript has been revised as per the reviewer’s suggestions. In particular, we have moved to the introduction the paragraph from the discussion section that argues that the previous literature examining adherence is primarily focused at physician knowledge deficits and awareness. We then outline how
patient social context factors may play a role in physician management decisions, and how this is something that is relatively less studied when examining adherence to CPGs. We have amended the introduction to better define what we mean by social context. We have also removed some technical terms, such as “choice conditioning” from the manuscript so as to make it easier to read.

2) Methods. Experienced and novice were well defined and justified, though not clear at first that five years of practice was after residency (which was later stated but should be in the opening of the methods.)

Response: The manuscript has been revised to clarify the study’s definition of an expert, as follows:

“We defined expert physicians as those who were residency trained in Emergency Medicine, and had at least five years of practice in EM after residency.”

3) The introduction of the survey instrument was clear but then became difficult to follow. Did each participant get one without a CV and then exactly the same one with a little tag that said, "now assume the patient is homeless"? That method would seem somewhat leading, which might be why they were given the cases in no particular order. At one point it sounds like no one got the same case in its non-CV and CV form, but then it is unclear why there are 2 survey instruments that had to be randomized. Basically, I think I finally understood how these cases were arranged and distributed after reading this multiple times, but it would benefit greatly by having both a figure showing the combination of cases making up each survey and how they were distributed to each set of physicians. It would also help to have a table listing just the basics of each case and CV rather than relying on just 2 examples in the text.

Response: Participants only saw one version of each case, which may or may not have included the context variable (CV). Participants randomized to a second group would see the opposite – that is, if participants in group 1 saw a case with CV, participants in group 2 would see the same case without the CV. The authors agree that the way this is described in the original manuscript is confusing. The manuscript has been revised to improve clarity of the design.
The basics of case construction are now described generally in the methods section, and a table outlining case specific information accompanies the manuscript (i.e. Table 1).

4) You collected a lot of demographic data not obviously related to the purpose or hypotheses of the study, and only reported on it in a demographic table. Just clarifying that it was collected to guide readers in knowing what population this study applies to helps guide the reader to how it will be reported. (i.e. you did not actually design the study to compare women and men or community vs. university physicians.)

Response: The demographic data was included for the purpose of describing the population as the reviewer states. We have removed the detail of how this data was collected from the “Data Acquisition” section, and have revised the “Analysis” section of the manuscript to emphasize the reviewer’s point, as follows:

“Data related to participant demographics and characteristics of the practice environment were used only to describe the population.”

5) It is also interesting to note that the physicians were not only "familiar" with a CPG but had to know the guideline by memory.

Response: The authors indeed assume that where the physicians indicated familiarity that this meant they had working knowledge of the CPG from memory. We have noted this in the “Limitations” section of the manuscript, as follows:

“While the physicians indicated familiarity with the aforementioned CPGs, this study assumed that they had functional knowledge of each CPG by memory.”

6) In the analysis, you do not mention a power calculation, which should have been done when comparing two groups to avoid Type II error. (Because the results were significant, it is evident that the sample size was large enough (though barely!), but if they had not been significant we would not have known if it were due to small size.) Given the very specific number recruited, I imagine you did some sort of calculation, but in any case it
would help to know how you came up with the 28 and 28 number. (Perhaps funding the honorarium was also an issue, a reality of research. If so, another reason to frame it as a pilot study as discussed below.)

Response: We calculated a sample size, a priori, to be approximately 20 participants in the expert group, and 20 participants in the inexperienced physician group. Our funding allowed for some additional participants.

While we would agree that a power calculation would be appropriate if we did not see an expected difference, in fact we found significant differences exactly where we anticipated them. In any case, a sample size calculation based on an alpha of 0.05 yields a sample size in the range of our study (if we guessed right on the differences). On that basis we see no reason to call it a pilot study. We have no basis to assume results would change with a larger sample.

The author’s by no means believe that this study is definitive on the issue of the role of social context in adherence to CPGs. However, we are not sure that this should be considered a pilot study, in that future studies should expand on this concept with different groups and cases, rather than repeat this analysis with a larger sample. We have noted the need for additional research using various specialties, and in different communities in the “Limitations” section.

7) In the data analysis you only mention using a chi-square to compare experienced vs. novice, though you actually also used it to compare CV with non-CV cases for each subgroup.

Response: The analysis compares CPG adherence in the CV and no CV conditions in the expert group, and then again in the inexperienced physician group. We then compare experts with inexperienced physicians in adherence to CPGs in the CV condition, and then again in the no CV condition. The “Analysis” section has been revised to clarify the analysis plan.
8) Results. Clear and well presented. Important that you noted a scan for individual outliers vs. general trends driving the percentages since each was responding to 4 different cases.

Discussion. Nice summary of results and literature review re-emphasizing that out of the possible reasons for CPG non-adherence, deliberate deviation based on patient circumstances and measured consideration by the physician is understudied compared to awareness and knowledge. (Though I would include at least a little of that in the introduction to clarify why you are doing a study in which the outcome seems obvious.)

Most of the discussion is a lot of information justifying the deviation from clinical guidelines and supporting your concerns that following guidelines like a cookbook is not responsive to many patients' needs, but those paragraphs could be condensed and should be consistent with your own arguments about more cost-efficient care. The Hughes and Boyd citations, for example, argue that multiple guidelines could result in extra treatment burden, while the CPGs you used in the study (with the possible exception of the CHADS score) tend to promote less use of expensive diagnostics or extra treatment, not more.

Response: The “Discussion” section has been revised to put less emphasis on how “social context” is understudied in the CPG adherence literature. Specifically, one paragraph has been moved to the “Introduction” so as to better justify the need for the current study, in light of the reviewer's suggestion (See our response to the reviewer's comment 1, above).

The paragraph outlining the Hughes and Boyd citations has been removed from the manuscript to make room for more discussion regarding the impact of the results on medical education, and to put less focus on the role of CPGs in reducing variation and improving efficiency, as this is not the intended focus of the manuscript. For that reason, the revised manuscript also condenses discussion on examining variation in medical practice.

9) You also need a more robust limitations section acknowledging the very small size (one of your main findings had a modest p-value with a CI that included 1) and
significant sampling bias in addition to the single specialty. I'm also not sure what management domains you were considering but then didn't power your study to find. The one citation on using case scenarios to represent actual practice is important to include, but still misses the #1 limitation of studies like this - that patient outcomes aren't measured or really even considered. You seem to imply that giving the remote-living nurse antibiotics for her URI is appropriate, but have no consideration for the physician who actually thinks of what might happen if she has an anaphylactic reaction or other significant adverse event while an hour or more away from the hospital. The point of population-based guidelines is that, if you had to choose applying them to everyone, regardless of "context variable" versus not having them at all, more people would be helped and resources saved than those who would be harmed. We have these because there is lots of evidence that reducing physician variation for these conditions does actually result in population-based improvements, so physician decisions to deviate from them must be really seriously considered. Is that what the experienced physicians were doing, or do they just consider themselves "smarter" than the CPG's? Or do they not understand evidence-based medicine or population-based guidelines? Perhaps those kind of questions are what you wanted to test with domains. (There are a lot more articles about why physicians don't follow guidelines than the 3-4 cited, and qualitative work might also strengthen your argument that your study is novel.)

Response: The reviewer raises a number of good points as to the limits of this study. Thus, the “Limitations” section has been expanded to include acknowledgment of these issues. Notably, the revised manuscript includes statements outlining that as outcomes cannot be and were not measured in this study, it is not clear whether the observed deviations in CPG adherence are indeed warranted (i.e. result in improved outcomes for patients), and thus, may or may not be justified. For example, the “Limitations” section now states:

“Finally, as this study used hypothetical cases, patient outcomes resulting from management decisions could not be measured. Thus, it is not known from this study if such deviations from CPGs would indeed result in improved patient outcomes.”

The authors agree that there are many more papers examining adherence than those cited here. For that reason, the manuscript cited 3 systematic reviews as examples of the literature in this area.
The authors agree that qualitative studies would indeed enhance our understanding of why physicians deviate from CPG recommendations in the presence of patient social context. However, such a study was beyond the scope of this particular experiment.

10) You acknowledge that physician decision-making (and by implication how we teach it to medical learners) is complex and that research should address much more than awareness or knowledge of guidelines and really get at the deeper processes and develop new ways to "guide" deviations from the guidelines. With that exact argument in mind, I would recommend re-framing this study as a pilot test that is examining a very specific element of deviation from guidelines. It adds to the literature by showing that in this select group of physicians who had almost universal knowledge and proper application of the guidelines, patient contingency and physician experience were both associated with deviations from them. It needs replication and bigger numbers and can't substitute for the other modalities of decision-making or actual patient outcomes, but as a pilot could also inform how to measure this modality in larger or more diverse studies. (These are things you said at various places in the manuscript but needs more focus consistent with its size and scope.)

Response: The authors agree that more study is needed in different specialties, with different cases, and in different communities. A larger sample is certainly desirable as well. The manuscript has been revised to emphasize this need (also see response to the reviewer’s comment 6, above).

11) Note also the "by implication how we teach it" above. In submitting to Academic Medicine, it is important to give some overt implications for education or academic institutions and not just rely on the fact that you happened to sample academic clinicians. The difference between experienced and novice participants gives you room to comment on that more specifically and you need to do so for this journal.

Response: The revised “Discussion” section presents further elaborates on how the results might impact medical education, as outlined in the included paragraph, as follows:
“Implications for Education

If patient context is viewed as a justifiable reason to deviate from CPGs, what is considered correct management for each patient may depend on individual characteristics related to the “context” of that patient.25,26 If so, how one learns to appropriately deviate from CPGs would require further study. The lack of complete agreement among physicians in this study as to when their recommended management decisions in the presence of the CV deviated from CPGs suggests that physicians react differently to the patient’s context. If and how physicians acquire information regarding social context in real clinical practice, and the relative importance they attach to it when making management decisions will require further study.”

Author Reply to Reviewer 2

Reviewer 2:
The authors describe an experimental study examining the effect of context on adherence to CPG in EM physicians and residents. Results indicate that, as expected, experienced physicians are less likely to adhere to CPG when context suggests not to compared to inexperienced physicians, and compared to cases where no context was present. While the results are in line with the hypotheses and the method and analyses seem adequate given the research question, I suggest rejection mainly because the authors fail to convince what the relevance and added value of this study is, both practically and theoretically. I will detail these and other comments below.

1. Particularly the theoretical value of this study remains unclear and limited. That is, what does this study learn us about medical expertise development apart from the less novel conclusion that experienced physicians incorporate context more easily? How are these data relevant to our knowledge of experts' mental flexibility (in their use of problem solving schemata)? To make the study more relevant, embedding it in the expertise literature would be needed. Still, I am uncertain whether that would improve the novelty sufficiently.

Response: The introduction of the manuscript has been revised to better clarify the added value of this study (see response to reviewer 1 comment 1). Specifically, previous
studies examining adherence to CPGs focus primarily on physician knowledge deficits or awareness of CPGs. However, we know from other literature that management decisions may be influenced by the patients “social” context. This study examined whether or not social context does influence (lack of) adherence to CPGs – a feature that is not well studied in the literature. The effect of expertise was a secondary outcome of the study. The authors acknowledge that more work is needed to understand what it is about experts that allows them to “justify” (to themselves) deviation from the CPG in light of the patient context, when compared to less experienced physicians. The authors believe that the revised manuscript better reflects these points.

Thus, the authors disagree that the study has nothing new to say. As pointed out in the revised introduction, while many studies have shown that experts depart from CPG’s this tends to be viewed as sub-optimal and a reflection of their lack of knowledge. The study clearly shows that the departures are by design, not ignorance and their knowledge is at the same level as residents.

2. Also, with regard to practical value, it is unclear why the research questions are considered important and relevant to examine. Is it problematical that novices are inflexible in their use of CPGS? Do they thereby cause errors? Do we need to adapt training/supervision because of that? Or, to what extent are these findings reassuring, given that by lack of experience, the novices choose the risk-reduced road consistently? Just as inexperienced drivers are taught all kinds of safety checks they will loose when gaining more experience.

Response: In view of the many exhortations in the literature to patient-centered care, and the concerns that individual needs of patients are not dealt with, much of which is also central to the Competency Based Education movement, the authors maintain that responding to patient individual needs is a laudable goal of medical education. Factors related to social context may important to patients, and thus, may present a justifiable reason for deviation from CPGs, which are context neutral. Furthermore, in some circumstances following the CPG may result in greater potential risk to the patient (e.g. not ordering a CT for a head injury in a patient who lives alone and cannot be monitored by a family member, or not ordering an ankle x-ray for a hockey player who may consider that a clean bill of health and will return to action). Thus, it is reasonable that a physician might be cognizant of these factors when ordering tests. The extent to which such factors influence (lack of) adherence is not well examined in the literature. To what extent deliberate non-adherence to CPGs in these circumstances results in better care is
not known from this study (this is indicated in the revised manuscript, see response to reviewer 1 comment 9). Thus, whether or not this is problematic for inexperienced physicians is not known. However, if it is believed that social context is an important feature of clinical practice, it is noteworthy that expert physicians are more adept at considering it in their management decisions. How this skill is learned is not clear, especially in comparison to content knowledge (that has the benefits of basic sciences, clinical trials, and CPGs for which to benchmark performance). Ignoring it in medical education provides opportunity for variation in practice, which many would argue is undesirable.

The “Discussion” section of the manuscript has been revised to outline how the results might impact medical education research (i.e. what information regarding context is important and how to integrate it with other clinical knowledge), and subsequently, medical education (See response to reviewer 1 comment 11).

3. From the abstract and introduction it is not clear that the crucial manipulation entails that the context actually suggests NOT to adhere to the CPG. The main research question driving the hypotheses should therefore be "Do physicians recognize when to depart from CPGs when context suggests to do so?" On the other hand, context that drives adherence to CPG is just as well possible? Why focus on situations in which context suggest deviance from CPG?

Response: The study was not intended to specifically promote non-adherence to CPGs, but rather, provide context features that might complicate strict adherence to CPGs and see if physicians respond to these features. The authors do not believe it is correct to reframe the study in its current form to an examination of whether physicians “recognize” when they depart from CPGs for two reasons: 1) the study makes no value judgment on if it is correct to deviate from CPGs in these circumstances, and 2) the study did not ask physicians if they recognized that they were deviating from CPGs (their familiarity with CPGs was asked after the completion of the case review). Description of the context variable design has been amended, and now states as follows:

“The CVs were designed to provide a reasonable basis for not following the CPG.”

As the focus of this study was examine if social context can influence non-adherence to CPGs (something overlooked in the literature as described in the revised introduction), introducing a social context that specifically and knowingly promotes adherence is
counterproductive to this purpose. The authors acknowledge that some contexts may indeed promote adherence, hence the challenge in incorporating patient context considerations into medical education. The “Introduction” has been revised to better explain our position:

“However, the examination of physicians’ deliberate non-adherence to CPGs as a result of patient context, in particular those personal or “non-medical” (i.e. social) in nature, receives comparatively less attention in the literature compared to that due to awareness or knowledge deficits.”

4. Also, if you are interested in studying to what extent expertise drives CPG adherence, a post test questionnaire examining how and why physicians use context (i.e., did novices process the context? What did they make of it?) would be needed to shed informative light on this issue.

Response: The reviewer raises an interesting idea. Unfortunately, a post test study was not performed due to 1) the desire to keep the participant’s time commitment low (this is a busy physician population), and 2) the online nature of the study did not allow for face-to-face interviews (for a qualitative component of the study). Thus, the specific question the reviewer raises is beyond the scope of the current study. With this issue in mind the manuscript has been revised to include the following statement:

“If and how physicians acquire information regarding social context in real clinical practice, and the relative importance they attach to it when making management decisions will require further study.”

5. Little detail is provided with regard to case construction, and case processing. Can you describe (by means of example cases) what the cases looked like? How they varied? And how context was inserted (and whether that varied between cases)? Can you describe how participants processed the cases? How cases were presented (sentence by sentence, completely?)? Was case processing self-paced?

Response: Detail regarding case construction has been added to the manuscript, specifically, the following statements:
“Each case described the presenting illness, its history, past medical history, current medications, social history, and results of the physical exam. A “context variable” (CV) that could be inserted into the social history was developed for each case. The CVs were designed to provide a reasonable basis for not following the CPG.”

“The presented information for each case appeared on a single page.”

“For each presented case, the participant was asked to indicate if he/she would order the diagnostic test or prescribe the treatment in question (yes/no). The participant could advance to the next case only after indicating their management decision. The survey was self-paced.”

We did not collect data regarding case processing, and thus cannot describe how the participants did this. Such a study would require a very different design, involving few subjects (further raising concerns about small samples, as expressed by reviewer 1), much transcription and perhaps strategies like eye-tracking.

6. With regard to the results presented, detail is missing too. What were differences in case processing? And you must have had some sort of performance measure? You would expect an expertise effect in performance at least. Present those analyses too.

Response: As described above, we did not collect data regarding how they processed the cases (i.e. how they acquired the information, where they focused their attention, how long it took to read the case, etc.). Thus, the manuscript does not include the analysis as per the reviewer’s interest. There are no independent measures of performance we can derive from these cases – for example, they were not diagnostic cases, and so we cannot examine accuracy. An examination of expert performance beyond the influence of social context is outside the scope of this particular study.

7. Residents who are familiar with CPGs and have been in EM practice for about two years can hardly be called novices, especially when compared to other studies in expertise development. Novices are typically those who have no or hardly any knowledge/experience in a domain and are typically beginning or advanced students. I
realize your expertise defining variable was experience with the CPG, but even at that level, these participants could better be termed 'inexperienced physicians'.

**Response:** There is no consensus in the literature about defining expert vs. novice. On the one hand, we have been taken to task by Ericsson because in medicine we tend to view “expert” as someone in practice for 5 or 10 years, and he views this as insufficient. In another large study of second year residents, we were challenged by the reviewer that they were not expert enough to justify conclusions about diagnostic reasoning. In the present context, we think that a) they are still only halfway through postgraduate education, and b) they are likely to be less expert at management than diagnosis, so we believe the label “novice” is justified. However at the end of the day, the label will remain arbitrary.

We did change the manuscript to clarify the study’s definition of expert and novice:

“We defined expert physicians as those who were residency trained in Emergency Medicine, and had at least five years of practice in EM after residency. We defined the novice EM physician group as those were enrolled in the first two years of a residency program in EM.”

8. There are a number of factors regarding the ordering of the tasks that probably were not crucial, but that should be corrected in future research. That is, order the cases randomly per list (unless you have crucial reasons why a specific order is needed, but then again, there is no order in case presentation in real practice either). And select cases randomly to fill a list per participant (in stead of using two fixed lists as you did now) to prevent list effects. I would advise to control for/test a list effect in the present study.

**Response:** The authors believe that using different cases for different subjects would create serious problems in interpretation because of case specificity, and would complicate analysis.

The authors agree with the reviewer that there is a possibility of an order effect. This has been noted as a limitation in the revised manuscript:
“The surveys used a fixed case order, creating the possibility of an order effect on management decisions.”

9. The final column of the final three rows doesn't present means (as the heading suggests) but sums.

Response: The revised table has added a space between the means and counts so as to avoid confusion. We have also added percentages to the table.