Author’s response to reviews

Title: What's in your wallet? A cluster randomized trial of the effects of showing comparative patient out-of-pocket costs on primary care prescribing for uncomplicated hypertension.

Authors:
Robyn Tamblyn (robyn.tamblyn@mcgill.ca)
Nancy Winslade (nancy.winslade@mcgill.ca)
Christina J. Qian (julingyi.qian@mail.mcgill.ca)
Teresa Moraga (teresa.moraga@mcgill.ca)
Allen Huang (allenhuang@toh.ca)

Version: 1 Date: 17 Nov 2017

Author’s response to reviews:
Reviewer #1

1) The reader needs more contextual information about:

   a) The rules about patient-responsible costs?

   b) Whether the insurance for eligible patients is the same or different for different individuals in regard to what is covered and how co-pays and deductibles are calculated. If there are differences, does the MOXXI calculation of OOP costs individualize its estimates for the computer alerts?

   c) The pre-existing frequency of other alerts presented to these clinicians and the extent to which they had alert-fatigue that would reduce the impact of a new one.

These are excellent points.

   a) We have added information about the prescription cost-sharing rules in the Context section of the manuscript.

   b) In the section describing the intervention, we have added how the out-of-pocket cost for each drug was calculated. This amount was personalized to the drug(s) selected. We did not
factor in the complexity of the monthly maximum contribution, as the out-of-pocket costs for a specific drug for patients who exceed the maximum contribution will depend in the order in which each prescription is filled. If, for example, the prescription for the antihypertensive is filled first, then our calculation of the out-of-pocket costs will be 100% accurate. However if a patient is filling prescriptions for 20 drugs, and the antihypertensive prescription is filled last, it is possible that the patient will have exceeded their monthly maximum and there will be no out-of-pocket costs.

c) In the section describing the control group, we have added information about the drug alert system. MOXXI provides physicians with the tools to select the alerts they want to see. As most physicians select to see only the most severe alerts (absolutely contraindicated) that represent approximately 5% of all alerts, alert fatigue should not have been an issue.

--------

2) It is one thing to say this is a test of an alert built into the computer, but you have told us nothing about the implementation of this new system. Since you are submitting this to IS, surely you should provide us with information about whether the technological changes for the trial were the only changes made or were there training, supervisory, incentives, champions, etc. involved in implementing these changes.

We have added a new section on physician training and support. MOXXI has been developed and continues to be modified to improve the clinical value for physicians. The expectation was that the out-of-pocket module would provide physicians with information they were interested in—their patient’s out-of-pocket costs.

--------

3) It sounds like individual physicians were randomized without regard for the likely clustering of their practice settings, clustering that might produce contamination or reinforcement of responses to the randomized study arms. Why isn’t that a consideration in blurring the very few differences you found between the two groups?

This is correct. Physicians were randomized, not clinics in order to obtain a better balance in physician and patient characteristics between the intervention and control groups. Physicians who are co-located in the same clinic and are part of the MOXXI program do not usually share patient management except for weekend call. It is plausible that an intervention physician may have discussed out-of-pocket costs with a control physician, and to the extent to which this may
have modified the behavior of control physicians, contamination will have reduced observed differences between the two groups. We have noted this as a limitation in the discussion.

--------

4) In the Background section, you describe efforts by governments and payers to introduce user fees to discourage the use of marginal therapies, with minimal effects. However, at least in the US, nearly all such efforts have lacked such discrimination between marginal and more cost-effective therapies. Except for restrictive formularies, they have generally relied on bureaucratic processes to make it harder to prescribe some kinds of meds.

We agree with these points. In the background we made the point that these efforts result in reductions in both essential and less essential therapy—i.e., they fail to discriminate between marginal and needed therapies and cause unintended harms by policy-induced non-adherence with essential medication.

--------

5) Since the trial is intended to test the impact of the alerts on physician and patient behaviors, shouldn’t the tables and analyses focus on the change in each outcome variable rather than simply on the adjusted risk ratio of the post-intervention results? Shouldn’t this be a differences in differences analysis presented in tables that show both pre and post values along with changes?

As we had to account for clustering in the analysis, we needed to use either generalized estimating equations or hierarchical models and modeling of risk differences are not yet available for these statistical packages. Moreover, we are adjusting for baseline differences in both patient and physician characteristics to provide an unbiased estimate of the effect of the intervention.

--------

6) Did the 76 included physicians constitute the entire eligible pool or are they a sample that might not be representative?

These physicians were the entire eligible pool.
7) The first sentence in the Results suggests that the average practice size is smaller than in the US -67,000 patients per 79 physicians or about 850 patients apiece. However, Table 1 shows it to be closer to 1300 - what is the difference? Incidentally, the practice size for the control group has an extra digit.

Thanks for pointing this out. We have clarified now in the first paragraph of the results section that the 67,000 are patients with hypertension. In table 1, the 1300 is the average practice size of all patients in the practice. We have corrected the decimal place for the control group.

--------

8) What was the pre-study adherence for both groups of patients, so we can see if the trial changed that?

Adherence can only be calculated for patients who were currently treated at the start of the trial. Among these patients, adherence improved in both control and intervention arms (from 65.3% to 70.7% in the control group, and from 68.3% to 72.1% in the intervention group). There appears to be no difference in improvement between the two groups.

--------

9) Finally, and most importantly, you never discuss the most striking results of the trial:

a) Even though there was a significant change in adjusted risk ratio for diuretics and only other anti-hypertensives among incident cases, there was no effect on either total costs or OOP costs.

b) Moreover, both arms of the trial appear to have decreased both total and OOP costs by about 25%. Is that a Hawthorne effect, or was that due to other changes in the environment during the trial? Shouldn't it be at least noted and discussed?

a) We inspected dispensed hypertensive medications among incident cases in the 12-month follow-up period. We found that, compared with the control group, a higher proportion of patients in the intervention group had medications from other classes of hypertensives added to their initial prescription of diuretics. This would explain why the statistically significant increase in prescribed diuretics in the intervention group had no effect on total costs or out-of-pocket costs. We have added a statement to this effect in the discussion.
b) We have clarified, in Table 1, that these are characteristics for all patients with hypertension (both complicated and uncomplicated) treated by the study physicians in the year before the start of the study.

Reviewer #2

1) The authors consistently employ the term "significant" in the statistical sense of the word. This is potentially confusing, and can even come across as non-sensical, e.g. in their conclusion-section, where they state that their intervention "resulted in a significant but modest impact. They should, in my opinion, add "statistical" before "significant", whenever they use the term in this sense. In the abstract I think they should write "had a modest impact" rather than "ha a significant impact".

Thanks for alerting us to this problem. We have gone through the abstract and manuscript and have indicated where we mean statistical significance.

2) In their description of the intervention software package, I get the impression that the physicians had to justify whenever they decided not to switch a patient over to a thiazide, but did not have to justify when the chose not to start a new (incident) patient on a thiazide. This is contrary to what I would have guessed. Is it correct?

Thanks for raising this point. We have clarified that physicians were asked to indicate why they did not choose a diuretic for both incident and prevalent patients.

3) The claim that physicians were blinded to the outcomes seems hard to believe. Didn't they realise that you were testing new software, and that the new software concerned promoting the use of thiazides? Even if this wasn't explicitly communicated to the participating doctors, they must have realised that this was the case?

We have added this caveat to the section on randomization and blinding.
4) "There was also no difference in the overall annual costs or out of pocket costs…" I think A) you should avoid using the phrase "no difference" (since you probably mean a statistically non-significant difference, which is not the same as "no difference"), and B) you should report the actual figures.

You are correct. We have changed the wording.

--------

5) From reading the Discussion-section I get the impression that you - in retrospect - probably should have user tested the software more extensively than you did. I find no description of the user testing. I think you should address this, as a minimum in the Discussion.

We agree! We have added this to the discussion.

--------

6) I see no mention of a study protocol. The emphasis you put on subgroup analyses makes me wonder to what extent these subgroups were pre-specified. I tend to think that the authors are putting a but more weight on the subgroup analyses than I find reasonable, but I guess it's OK if they were pre-specified AND all pre-specified sub groups are reported in the manuscript. If they were not pre-specified, I would want you to tone down the subgroup results, quite a lot.

The study protocol registration number is cited at the bottom of the abstract. Subgroup analysis was planned as we find that interventions work better in some groups compared to others and this is a means of gaining more knowledge on how to better personalize interventions for both patients and clinicians. The list of potential modifiers that were examined is provided in the methods and results sections. We assessed whether the intervention was modified by patient characteristics such as age, co-morbidity, level of co-payment required by the drug insurance plan, estimated household income, and the patient’s co-payment plan. We also fit two-way interaction terms between the intervention and a number of physician characteristics, including years in practice, sex, practice volume, average costs of prescribed anti-hypertensives, and propensity to prescribe diuretics.