Author's response to reviews

Title: A quasi-experimental test of an intervention to increase the use of thiazide-based treatment regimens for people with hypertension

Authors:

Carol M Ashton MD MPH (cashton@uab.edu)
Rebecca J Beyth MD MS (rebecca_j.beyth@med.va.gov)
Tracie C Collins MD MPH (tcollins@bcm.tmc.edu)
Howard S Gordon MD (gordon/howard@va.gov)
Paul M Haidet MD MPH (phaidet@bcm.tmc.edu)
Myrna J Khan PhD (mikhan@bcm.tmc.edu)
Barbara Kimmel MS MSc (barbara.kimmel@med.va.gov)
Anna L Kolpakchi MD (kolpakchi.annal@med.va.gov)
Michael L Johnson PhD (mjohson@bcm.tmc.edu)
Lee B Lu MD (lblu@bcm.tmc.edu)
Aanand D Naik MD MPH (anaik@BCM.tmc.edu)
Laura A Petersen MD MPH (laurap@bcm.tmc.edu)
Hardeep Singh MD MPH (hardeeps@bcm.tmc.edu)
Elizabeth Stanberry PharmD (elizabeth.stanberry@med.va.gov)
Annette Walder MS (awalder@bcm.tmc.edu)
Nelda P Wray MD MPH (nwray@bcm.tmc.edu)

Version: 2 Date: 30 October 2006

Author's response to reviews: see over
A Quasi-Experimental Test of an Intervention to Increase the Use of Thiazide-Based Treatment Regimens for People with Hypertension (Ashton et al)

Response to Reviewers’ Comments

Reviewer Wessell

Major Compulsory Revisions

1. Limitation: no accounting for confounders; repeat analysis with confounders or explicitly state why this was not done.
We now present in the paper stratified analyses by diabetes status and age. Please see our response to the concern #2 of the statistical reviewer for a full explanation of our other efforts in this regard.

2. Describe the simultaneous QI intervention in thiazide use that was taking place in the control group (PrimeCare).
The relevant paragraph in the Methods has been revised to read: “After our intervention period ended, we learned that a contemporaneous quality-improvement project aimed at increased thiazide prescriptions was occurring in PrimeCare, consisting of distribution of copies of the ALLHAT main results publication, several lectures on ALLHAT’s findings, and the introduction of an electronic reminder to consider a thiazide diuretic into the medical record of every patient whose blood pressure was not controlled at the time of the visit and was not prescribed a diuretic.”

3. Move data on age and medical conditions data to Results Section; move timeframe descriptions to Methods Section.
Done as requested.

4. In the Discussion, discuss as a strength of the study the high prevalence of minority patients, and describe how your study population compares with ALLHAT’s.
Done as requested in paragraph 5 of the Discussion.

5. Strengthen the conclusions statement in the Discussion. Address benefits of EMR first then discuss and expand upon limitations of the Rogers model. State what other limitations and barriers to improvement were encountered.
Done as requested; please see last paragraph of Discussion.

Minor Essential Revisions and Discretionary Revisions

1. Revise so that informal language (words such as dazzling, boost, and dipped) is removed.
Done throughout as requested.

2. Make verb tense consistent throughout.
Done as requested.

3. Add numbers and percentages to text of Results, not just in Table.
Done as requested.

4. In Table 2, consider adding a column showing how each team’s charge relates to the Rogers model.
Done as requested.
5. In Results, state whether there were any non-physician providers in PrimeCare. This is stated in the second paragraph of the Methods.

6. Figures were blurry in printed form. We have redone them and they are clear now. We apologize for the trouble of the reviewers.
Reviewer van der Weijden

Major Compulsory Revisions

1.a) On pg 9 it says that the implementation of ALLHAT findings has been insubstantial. State why it was insubstantial.
We now provide in paragraph 2 of the Introduction an additional and more recent reference for readers documenting insubstantial (though small post-ALLHAT increases) in thiazide prescribing. While we can speculate why thiazide prescribing is not as widespread as the literature indicates it should be (we think it has to do with marketing of other, more expensive products by drug companies), there is no proof, and we did not think it was appropriate for us to include our speculations in the paper.

1.b) Rogers diffusion model was used, but all the factors in the model seem to refer to reasons for professionals to change while the real barriers might be more organizational or patient related.
The paper’s last paragraph has been revised to clarify the reason we chose the Rogers model originally and why, once the implementation intervention had been designed and implemented, it proved unhelpful. Sustaining such an intervention requires sustained organizational change and a suitable model for organizational change. In accordance with what the reviewer is saying, we agree that the Rogers model was appropriate for communicating the ALLHAT innovation to members of the clinical microsystem, but would not be appropriate to use to sustain organizational change.

1.c) I do not understand the sentence “We did not explicitly address compatibility because cost-effective care is a cultural cornerstone in the VA medical care system.”
We added to this sentence the qualifier “publicly-funded” VA medical care system as further explanation. The point is that the VA system is the closest the US comes to an equal access health care system funded by public sources. In health care delivery systems funded by public tax dollars, cost containment is a cornerstone of the system. This is not the case in fee-for-service arrangements.

2. How was the target population defined exactly?
We used the approach that we developed for another study and that we cite (Johnson et al, Am J Managed Care 2004;10:926-932). That paper presents a lengthy description of the algorithms we applied to the electronic medical records to identify the target population and their characteristics.
To provide more information for the reader of the new paper we now include a sentence describing the derivation of the target population in paragraph 4 of the Methods.

3. The potential bias stemming from the nonrandomized design must be more discussed. Instead of saying “the direction of the potential bias caused by differences in features of professionals is harder to predict”, I would say, “it can be argued that the difference in professionals might have caused the difference in change between groups instead of the intervention.”
We have revised this sentence to read, “However, features of practitioners were also different between the groups, and the extent to which differences in their behavior explain the consistently higher proportion of GMS patients on thiazides and meeting blood pressure goals cannot be determined from this study design.” Respectfully, we disagree with the reviewer that differences between the professionals might have caused the difference in change between the groups; we think it may explain the consistently lower use of thiazides and BP goal attainment in the PrimeCare group.

Minor Essential Revisions
1. In the Abstract change the language from “...were taking thiazides” to “...were prescribed thiazides.”
   Changed as requested.

2. In the Discussion acknowledge that you assumed that no under-diagnosis of hypertensives and no over-diagnosis of hypertensives were operating.
   We now include a sentence in paragraph 4 of the Methods stating that “The misclassification rate of these algorithms has not been empirically assessed”.

3. (p20) The sentence “our intervention appears to have led to better blood pressure control rates, regardless of medication regimen” is not clear.
   We have revised this sentence to read, “More importantly for patient outcomes, our intervention appears to have led to better blood pressure control rates, regardless of whether a thiazide was prescribed or not.”

4. Print versions of figures can hardly be read.
   We have redone them and they are clear now. We apologize for the trouble of the reviewers.

5. You might be over-interpreting your data when you say that “the percentage of patients who achieved BP control was 51%, only slightly lower than what was achieved in ALLHAT.” Were your patients comparable to those included in ALLHAT?
   Except for gender, they were comparable. We now point out in the Discussion that “Because this study was performed in a VA setting, almost all patients were men; however, the proportion of black patients (29.5%) and patients with diabetes (40.5%) was roughly similar to ALLHAT’s (35% black, 36% diabetic).” We have revised the sentence the reviewer questioned to read “The percentage of GMS patients who achieved goal blood pressures in this project—51.6%—approaches the 55.2% achieved at one year in the ALLHAT.”

Discretionary Revisions
1. Some information in the paper seems unnecessary (list of 5 things plus Table 2)
   Item #1 info on ALLHAT: We believe the average reader needs the background information on ALLHAT (and comments from the other two reviewers suggest this to be true) and so have decided to retain it.

   Item #2 value of last sentence, pg 9: The value of the sentence telling readers that the NHLBI has decided to fund a large implementation study to drive ALLHAT results into practice is that it helps readers develop a perspective on the value of our study.

   Item #3 info on VA: We have retained the info on VA because many readers do not know much about the VA and the contextual information for the implementation intervention is very important.

   Item #4 info on GMS physicians: We retained it to help readers understand differences between the two study groups in this nonrandomized study.

   Item #4 info on research initiative: We deleted it.

   Table 2 can be left out: We decided to retain it because it helps readers understand the facets of this complex intervention
2. No adjustment has been made for baseline differences. Consider multivariate analysis. Was such a large sample necessary? Were statistics necessary? Did it increase the risk of finding statistically significant but clinically irrelevant findings?

RE adjustment for baseline differences, please see our lengthy response to concern #2 of the statistical reviewer. Re the sample, paragraph 4 of the Methods now describes how we identified the target population. Because we used previously-developed algorithms on the electronic data warehouse, there was no advantage to us to use a partial random sample. Regarding the risk of detecting statistically significant but clinically irrelevant findings, we acknowledge in the paper that the intervention had modest effects, and we present for the reader’s assessment the raw data on the magnitude of increases in thiazide prescriptions and BP goal attainment rates.
1. Given the apparent multiple evaluations of the data through the monthly determination of the cusum or the proportions, isn’t there a problem with Type 1 error rates in the analysis? I realize in quality control applications where these techniques are typically used this is not really a concern because of the use made of the charts, but there are inferential determinations being made here...

The control limits used in the statistical control charts are confidence intervals, mean ± 3 SE. So when a point is within the control limits, the null hypothesis (process is “in control” or in our case not changing) is not rejected and when a point is outside the control limits, the null hypothesis is rejected and the alternative hypothesis (process is “out of control” or in our case, changing) is accepted. Using 3 x SE is equivalent to testing using a type I error probability alpha = 0.0027, so this level should be an adequate level to correct for multiple testing. Our type II error rate (an out-of-control point falling within the control limits) will tend to be higher, but that is conservative.

The CUSUM (change point) procedure is less clear about how it controls for multiple testing. The results are given in terms of confidence intervals along with a confidence level. Another level is given which indicates the importance of the change. A level 1 change or primary is the first change(s) detected. In this analysis, we are reporting only primary changes. Any other changes are detected by subsequent passes through the data. Using only a level 1 change controls for multiple testing.

2. A fundamental flaw of such an approach is an inability to handle baseline covariates reflecting differences between the groups. Because of the relatively small numerical intervention effects seen, the authors should consider an adjusted analysis in order to buttress their results. They might consider a repeated measures analysis with covariate adjustment.

In the ALLHAT study, certain attributes were associated with a lower likelihood of being at BP control at 36 months. These included being male (almost all our patients are male), diabetic, older, and black. The proportions of black patients was roughly the same in GM and PC, but GM had older patients and more diabetics. Therefore, the important baseline covariates for our analyses are the presence/absence of diabetes, and age, which we dichotomized as >65 or <65.

To respond to the reviewer’s direction, we conducted extensive multivariable analyses with and without lag periods. We examined segmented regression analyses including pre-intervention, intervention, and post-intervention periods, and Poisson regressions including adjustments for possible correlations between subjects using the SAS autoreg procedure and Generalized Estimating Equations, respectively. These analyses did not shed any light on the potential influence of confounders, for the following reasons: the data themselves show a non-linear, almost cyclical nature (this can be seen in the control charts), and these types of linear and log-linear modeling approaches do not fit well; secondly, the aggregate nature of the data restricted us to only 33 observations, which does not lend itself well to regression with covariables.

In consonance with the context of the study and our desire to present findings in a form accessible to most physicians and health care administrators, we conducted simple stratified analyses, which we now present in tabular form in the paper.

3. The authors argue that clustering should not be an issue in at least one of the study groups, but I would like to see an evaluation of their data that actually supports their ignoring of this potential effect.
Clustering is a potential issue only for the pre-post analyses of proportions between GM and PC. We are unable to test for it because the datasets do not contain an identifier linking provider with
patient. To help readers understand how it might affect the results, we now state in the Methods, “Clustering effects are unlikely in GMS panels: the large number of trainees and their rotation schedule led to a large number of different providers, small numbers of patients per panel, and short exposure of panels to specific physicians. Clustering might have been present in PrimeCare, and we did not adjust the pre-post analyses for it. To the extent such clustering was present, it would have led to a smaller standard error for the pooled proportions, which would have increased the value of the test statistic (z score) and lowered its p value.”

4. *The comparison of proportions pre and post seems over-simplistic given the time series nature of the data. An examination of Figures 5 and 6 do not suggest a particular noticeable difference in the patterns of the BP goal outcome post-intervention. Does a time-dependent analysis demonstrate any significant difference?*

We agree that a pre-post analysis alone might be over-simplistic, but we felt that this analysis, in conjunction with the control charts and the CUSUM analyses, are understandable by a broad array of readers and addresses the primary questions regarding whether change occurred over time. The change-point analysis further addresses the questions of when change occurred. Time-dependent analyses did not seem appropriate given our goal to maintain an analysis that could be exportable and usable by hospital administrators.

5. *The large sample sizes probably guarantee such findings anyway. The authors argue that the small differences seen are clinically important, but were there any sample size and power calculations performed pre-study? Such a discussion belongs in the Methods section.*

There were no sample size or power calculations performed prior to the study. From our prior work, we knew that the prevalence of hypertension in our study population would be greater than 50%, and that the large sample might lead to the detection of differences that while statistically significant were not clinically so. We now state in the Methods, “We developed the computer algorithms for defining the target population and measuring key variables. We used all patients who met the conditions in our defining algorithms and did not perform any sample size or power calculations beforehand. The target population consisted of patients who had…”