Author's response to reviews

Title: The Diagnostic Yield of the first episode of a Periodic Health Evaluation

Authors:

Cindy A Kermott (kermott.cindy@mayo.edu)
Carol S Kuhle (kuhle.carol@mayo.edu)
Stephanie S Faubion (faubion.stephanie@mayo.edu)
Ruth E Johnson (johnsonmd.ruth@mayo.edu)
Donald D Hensrud (hensrud.donald@mayo.edu)
Mohammad H Murad (murad.mohammad@mayo.edu)

Version: 2 Date: 7 March 2012

Author's response to reviews: see over
Dear Editor,
We thank you and the reviewers for taking time to review our manuscript. We made all the possible changes as recommended and we hope this version is more suitable for the readership.

Cindy A. Kermott and coauthors

**Editor's Comments:**

1. Please specify name of ethics committee who approved the study

   Mayo Clinic Institutional Board review approved the study. This statement added to the manuscript.

2. Kindly include a statement on informed consent

   A statement added regarding the informed consent process: 
   “Only patients who had consented to have their medical record data used for research purposes were included.”

3. Please move funding from title page to Acknowledgment

   This is done.

**Associate Editor's Comments:**

"Please address comments and concerns raised by the 2 reviewers."

This is done; see below.

1. Title page

   This should contain; Title, Author list, Affiliations (department names, institution name, street name, city, zip code, country), email addresses. The author list and email addresses must be identical in the manuscript file and on the submission system, and it must be clear which affiliation pertains to each author.

   This is done as requested.

2. Tables

   Please note that we are unable to display vertical lines or text within tables, no display merged cells: please re-layout your table without these elements. Tables should be formatted using the Table tool in your word processor. Please ensure
Reviewer: Nynke Van Dijk  
Reviewer’s report: 
In this study the authors aim to identify the diagnostic yield of a first periodic health evaluation. Although any part of additional information on the value of preventive/ screening services is useful I do have some concerns.

Thank you so much for your comments and I have responded to the best of my abilities as noted in red highlight below. Should you have further need for additional information, including raw data, please feel free to inquire.

Major compulsory revisions.

My main concern with this study is that the periodic health evaluation as described in this study is a very extensive evaluation with a wide variety in clinical disorders as outcomes. The current conclusion of the paper is that the full (2 hour plus) evaluation is needed to obtain the diagnostic yield presented. Although I agree that for instance Obesity and nicotine abuse are important health problems, they do not require the full PHE as described, but a few, relatively simple questions and measurements can be used to identify these problems. These diagnoses therefore do not require the full PHE. The important question therefore to me is not what the diagnostic yield of the PHE is, but also which components of the PHE are important in making these diagnoses. This would result in a more critical evaluation of this time/ money consuming effort. The PHE consisted of additional testing, which was, fortunately, partly based on the symptoms, signs and specific situation of the patient. Without knowing which and how many tests were done, we remain unaware of the testing needed to obtain these numbers of diagnoses. The judgment whether the evaluation is worth the investments, in my opinion a major outcome in making decisions on preventive healthcare, remains impossible. Could you describe the additional testing required?

In general we are in agreement with the reviewer’s point of view. This is, indeed, a descriptive study of an extensive evaluation encompassing a wide variety of clinical disorders as outcomes. Hence, the results apply best to these types of evaluations. These evaluations are not rare and occur in various settings, whenever prevention and disease management are done simultaneously. Our goal was to describe the possible clinical outcomes or diagnoses of a first PHE of a new patient in this context. Our study had not focused on efficiency or cost and we do not claim that this approach (detailed and comprehensive PHE) is the best
approach. We did not compare this approach and we simply describe its
diagnostic yield.

The reviewer also astutely points out that it will be very useful to know what
components of the PHE were most valuable. We believe that each diagnostic
entity would be associated with its own unique strategy for identification. Hence,
we are unable to determine such components.

In the revised manuscript, we added two statements consistent with reviewers’
point of view, that is, limitations to our data in terms of applicability and non
comparative nature; and the inability to determine the value of components within
the evaluation.

Last, the frequency with which these PHEs should take place is under debate.
Although this was the first comprehensive PHE for the included subjects, the
information on earlier participation in less extensive preventive programs as
information on their health status is missing. This could however provide an
important insight in the additional value of this program.

In our program, patients may return annually, biannually, every three years or
more and based on their choice. It often reflects geography as well, as those at
a distance return less frequently. We selected the first visit as a starting point,
and excluded any known, older diagnoses in our analysis. (Part of the value of
the PHE, not expounded upon in this study, is the management of known chronic
diseases as is also provided.) Many have local primary care physicians. In a
concluding letter to our patients we send a duplicate copy of their entire medical
record to them and encourage them to share these results with their local
physician to ensure continuity and optimal care coordination.

We did not collect information about earlier participation in less extensive
preventive programs. However, in a previous study, a 6889 patient sample from
the same cohort (Shippee et al), we demonstrated fair, although variable,
adherence to preventive recommendations (ranging 62%-91%). In the revised
manuscript, we added this information from the Shippee study and brought the
attention to the issue of PHE frequency highlighted by the reviewer,

Minor essential revisions.
Introduction:
“Experts have advocated for the delivery of preventive services in the context of
ongoing
clinical care.3-5 In fact, the delivery of preventive services and screening tests in the
context of a PHE has been used as a rationale for this evaluation.6,7” This part is
not clear for non-US readers.

The two sentences were re-worded for better clarity.
Methods:
- Could you describe in more detail based on which information data-extraction took place? Were only diagnoses as coded by the physicians included? How did you make sure diagnoses did not already exist (i.e. class III obesity should have been identified earlier?)

Nurse abstractors were trained to extract patient diagnoses from the clinical notes (i.e, documentation of the treating physician in the medical chart). We did not use (ICD codes), given they provide less than ideal detail. The principal investigator of the study trained the nurses and conducted several quality assurance checks and reviewed several samples of their work.

Content was further analyzed to make the determination of whether the diagnosis was old or new (this had been the more laborious part of the abstraction process). We used the following framework:
- If a symptom (patient-prompted) or clinical sign (exam finding) led to a conclusive diagnosis, the diagnosis is listed as new.
- If the clinical note contained wording that suggested a pre-existing diagnoses (terms like “personal history of”, “known”, etc.), the diagnosis was listed as old.
- If related symptoms are mentioned on the forms provided by patients in preparation for their visit (usually patients fill several forms that include information on their present illness or past medical history)
- Typically, but not always, patients will bring in their outside records for the first visit; if these records indicated an existing diagnoses, these had been characterized as old and known.

All old diagnoses were excluded, and all data presented in the manuscript is about *new diagnoses. In the revised manuscript, we clarified the methodology and added more details.

In the revised manuscript, we clarified the methods section.

- Could you describe in a little more detail which diagnosis were and were not included?

Our paper focused on new diagnoses. The diagnoses that were excluded were old (existing diagnoses) and trivial diagnoses, both risk and disease. The methods section is clarified and detail provided below as well:

Severity 1: Cerumen impaction, Cherry angioma, Excessive caffeine, Menopausal symptoms, Menes irregularity, Nevus, Perimenopause, Rosacea, Seborrheic keratosis, Skin tags, Telangiectasias
Severity 2: Anemia NOS, Benign prostatic hypertrophy, Diet inappropriate Dyspareunia, Eczema, Hearing loss symmetric high frequency, Hemorrhoids, Hypercalcemia, Neck and shoulder pain, Pre-hypertension, Stress.

This information has been placed in the Results section.

- What was the size of the random sample on which the ICC was based?

Random numbers table utilized to select 44 diagnoses for ICC calculation; stratified random sampling of eleven each of severity 1 through 4. The methods section is clarified.

- Please add ethical considerations. Patients were asked for their consent I presume?

Yes, patient consented and the study was approved by the IRB of Mayo Clinic Rochester. We have clarified wording within the manuscript.

IRB protocol 1935-05.

Results:
- From the paper I do not understand why the description of the population as described in the methods (executive health cohort) differs from that in the results section. Please fully describe the sample used for this study (i.e. including gender).

That sentence refers to the whole cohort, from which the current sample is derived. We clarified that sentence and added table 1 that describes patient characteristics of the current sample.

- Of the 428 new diagnoses, 82% (350/428) were in men and 18% (78/428) in females. Please relate to the size of the male and female population. The diagnostic yield in women seems to be very low. Could any other differences, for instance in age groups, be identified? This is worth considering when advising patients to participate in a PHE. By using multivariate models predicting those patients likely to obtain a clinically significant diagnosis, possibly a selection of patients who optimally benefit from this kind of testing could be made.

We thank reviewer for this excellent suggestion. We estimated the odds ratio (OR) and 95% CI of having at least one diagnosis using logistic regression with three explanatory variables, sex, age, and number of concerns or complaints listed on the admission intake form filled by patients. We founds that men (odds ratio 2.67; 95% CI, 1.76, 4.03) and those with multiple complaints at presentation (odds ratio 1.12; 95% CI, 1.05, 1.19) were more likely to receive a clinically relevant diagnosis at the conclusion of the visit. Age was not a predictor in this
cohort (perhaps due to small spread of data points). This was an excellent suggestion by reviewer and the manuscript is revised to reflect these findings in the abstract, methods, results and discussion sections.

- Page 9 – last paragraph: the sample used for this study, as they are employed, are probably also in better health than the general population, as most ill people do not have the ability to work. How would this affect your results?

Yes, it will likely affect the diagnosis. This is an external validity point raised by reviewer. In the revised manuscript, we highlighted this issue in the discussion section and we thank the reviewer.
Reviewer’s report
Reviewer: Amy Linsky
Reviewer’s report:

Thank you so much for your comments and I have responded to the best of my abilities as noted in red highlight below. Should you have further need for additional information, including raw data, please feel free to inquire.

Major Compulsory Revision
General comment - One of my biggest concerns with this manuscript is that they reference the periodic health evaluation, which my perception of - and the background alludes to - is a recurring annual exam without regard to acute complaints. However, the cohort that the study analyzes is the first exam for individuals presenting to an executive health physical. The first exam for any patient to any practice seems conceptually different than analyzing a return visit for a periodic health evaluation. Related is the fact that the cohort is from an executive health clinic, which the authors themselves discuss as a limitation in the discussion section.

We agree with reviewer although this first evaluation conducted in our program is likely one in a series of evaluations in a particular patients’ adult lifetime; hence, the PHE implications. Nevertheless, the reviewer points out an excellent point. We made revisions throughout to always emphasize the adjective “first” with PHE and to make sure that our results are not extrapolated to a “periodic” evaluation.

Specific comments:
Intro Section
Third paragraph - is there a citation regarding patients “saving” up complaints for their PHE?

No specific citation, just our own clinical observation when they pull out their yellow note pads at the time of their first visit!

Fourth paragraph - The last two sentences of the Intro seem more appropriate for a discussion and don’t seem to fit here.

Changes were made as requested. Sentences moved to discussion section.

Methods Section
First paragraph - As mentioned previously, I think there needs to be a justification for using the first exam to a clinical site; do these patients have other health care providers who they have been seeing regularly? Are the characteristics of the executive health cohort the same as for the cohort used for this study? Are the cohorts the same? I would like to see a table with baseline characteristics rather than just a reference; who are the participants in this study? What is there health status previously? What type/amount of care have they received prior to this
visit? Some of these answers seem better suited for the results section.

Second paragraph – is the conclusion of the medical episode after the first visit or after the second time meeting with the patient? How was it determined if a diagnosis was “new?” Was a diagnosis considered “old” if the patient reported it? Or did they look for previous medical record documentation, either at Mayo or elsewhere?

These are all excellent recommendations by reviewer 2 and many of them have been raised by reviewer 1. We made all the required revisions with additional detailed answers to reviewer 2. Basically, the revised manuscript contains table 1 that describes this sample. We clarified in the methods section that the Shippee paper describes the whole cohort (from which this sample is drawn).

We clarified how diagnoses were designated to old vs new. This had been the more laborious part of the abstraction process. If a symptom (patient-prompted) or clinical sign (exam, finding) led to a conclusive diagnosis, it had been listed as new. Wording within the document (“personal history of”, “known”, other wording in the HPI (history of present illness), listed in the Past Medical/Surgical section, or in a form (PFH—past and family history scanned form provided by patient before start of visit) had been utilized to make this specific distinction. Trained nurse abstractors had completed that piece. Given that this was a new visit, no Mayo records would exist. Typically, but not always, patients will bring in their outside records for the first visit; certainly these would be utilized to craft a more accurate Past Medical/Surgical section that had assisted us.

We clarified that we could not assess the quantity and quality of prior medical care, which is a limitation of our study. We also clarified that this “first” evaluation in the program consists of 60 minutes initial visit, followed by a second visit (usually averages 15 minutes and often is done the following day after all testing results are available). Most, if not all, diagnoses had been recorded after the conclusion of the second visit (i.e., the end of this episode of care). This is an important point because it demonstrates a unique structure of this healthcare delivery method that may decrease external validity to other models with different structures. Nevertheless, we think it is important to describe the expected outcomes of this unique model because there are other similar programs (many exist in the US) and also because this model may be beneficial in certain settings or patient populations.

In the revised manuscript, all these details are clarified.

Third paragraph – what scale of clinical importance was used? Is there a list of what was considered trivial (either a table, or reference, or online appendix)? Is everything that is not trivial thereby considered to be “clinically important?”
There is no known and validated scale that categorizes the importance of diagnoses. We used clinicians’ judgment (with good inter-rater agreement) to exclude trivial diagnoses. In the revised manuscript, we provide more examples of these excluded diagnoses.

Is intraclass correlation the right test, or is there a need for kappa?

We did not use kappa because we had more than 2 raters (weighted kappa can be estimated in this setting but the intraclass correlation test is more appropriate and has more power). ICC can be interpreted similarly to Kappa (>=0.75 is consistent with good agreement, 0.40-0.75 is good agreement, > 0.4 is poor agreement). We added this last sentence to the methods section and we thank reviewer.

How was a random sample of diagnoses chosen? How many were selected?

A random number table was used to select 44 diagnoses used to estimate agreement. Manuscript is revised to clarify.

Fourth paragraph – for non-patient prompted, is more detail available (such as from history, physical, labs, radiology)?

This level of detail had not been specifically abstracted but often it is a composite of findings where myriad detail is pulled together like a puzzle. It would have also been a difficult task even for trained nurse abstractors. Overall, the value and diagnostic yield is enhanced to a substantial degree when all available data is at hand, changing the positive and negative predictive value accordingly. As an interesting second study, it would be good to get a sense of what detail was required to make a particular diagnosis, in order to improve efficiency. Our study goal was to demonstrate diagnostic value, instead.

Results Section
First paragraph – Who is the study population? How many total diagnoses did each patient have? I could see that if a patient had a lot of pre-existing diagnoses, there would be less opportunity to make a new diagnosis. The 428 “clinically important” diagnoses are for how many patients (I see it in paragraph 2, but would like to see it sooner). Second paragraph – Are the 255 patients with new diagnoses different than the 246 without new diagnoses? describe the 491 patients.

We apologize for lack of clarity. All the questions posed by reviewer should hopefully now be clarified in the revised manuscript. We added a table 1 that describes the characteristics of patients that had diagnoses and those that did not. Data presentation is now simplified and more clear. We also clarified the reference to the characteristics of the whole cohort (from which the sample of the current study is derived).
Not sure that mean diagnoses per patient across all patients (even those with zero diagnoses) is helpful.

The mean of 0.9 diagnoses per patient may be helpful to a policy make or someone who is planning to establish a program that offers PHEs. In our program, we found it helpful to know that in a given population of particular size we would expect a certain number of diagnoses that may require downstream services such as testing and consultations. Please note that we also provide a mean diagnoses per patient excluding those without any diagnoses (which is 1.7). Hence, both rates are provided.

What does continuously enrolled mean?
We removed the word continuously enrolled (which intended to mean a sample of continuous patients presenting).

Third paragraph – as mentioned in the Methods section, is there any more information about how the non-patient prompted diagnoses were made? Please see detail listed under Methods fourth paragraph.

Discussion Section
First paragraph – is this prospective? Or was a retrospective chart review conducted?

The protocol of the study was established before patients were seen by physicians (before the PHE). However, data were collected retrospectively (i.e., we waited until patients were evaluated, had testing and had diagnoses issued by physicians; then abstracted data from the medical records). We decided to change to describe the study design to retrospective.

Second paragraph – This first sentence is long and somewhat difficult to understand. The concept of unhealthy lifestyle behaviors/habits, while I agree is important, seems a bit tangential. Further, these numbers should be presented in the results section.

We thank reviewer for the suggestions. Sentence re-worded for better clarity.

Sixth paragraph – it gets confusing with the references to USPSTF recommendations – I think it would be helpful to either reference the recommendations that you used more explicitly, or potentially focus just on the value of the PHE for diagnosing USPSTF related diagnoses. That is, make a case that the PHE is (or is not) a valuable service to provide the recommendations of the USPSTF. It may be more objective than using “clinically significant” diagnoses.
We agree with reviewer and this sentence was clarified.

Seventh paragraph – I agree that this is important, and I was curious in the methods about the scale used in this study. However, this isn’t really the study objective and I’m not sure it helps the discussion. Consider deleting it, and maybe putting something in the methods like, “Given a lack of valid and reliable scales, we chose to determine significance in the following way…”

We agree with reviewer and this was revised as recommended.

Conclusion
Do you mean to say viable? Or are you trying to say it's valuable? If the latter, not sure they can make that conclusion without any outcomes data.

Conclusion revised and the word viable is corrected.

Table
For prevalence, I would like to see the n (%), since the n’s appear to be small. I would consider just showing patient prompted since the “/” makes me think it should be a ratio. Or show the inverse, how many were NOT patient prompted, since that may make a stronger case for the value of the PHE. This is a really long table. Is it possible to highlight what you think are important findings?

Thank you for this good suggestion. We added n (%) to the first column and only presented the non patient prompted in the second column. We could shorten the table if editor requests to do so although with the journal being online, there is no compelling reason to do so.

Minor Essential Revisions
Discussion – paragraph 4 – “…2.14 diagnosEs”

Thank you, typo corrected.