Author’s response to reviews

Title: Correlating pharmaceutical data with a national health survey as a proxy for estimating rural population health

Authors:

Ronald E Cossman (rcossman@ssrc.msstate.edu)
Jeralynn S Cossman (cossman@soc.msstate.edu)
Wesley L James (Wes.James@memphis.edu)
Troy Blanchard (troy@lsu.edu)
Richard K Thomas (richardkthomas@comcast.net)
Louis G Pol (lpol@mail.unomaha.edu)
Arthur G Cosby (arthur.cosby@ssrc.msstate.edu)

Version: 3 Date: 28 May 2010

Author’s response to reviews: see over
Title: Correlating pharmaceutical data with a national health survey as a proxy for estimating rural population health

Note: We have responded in italics to individual comments. Additionally, IMS Health (the Rx data provider) comments are set off with parentheses.

1. The editors discussed your manuscript in their monthly editorial meeting. They find the idea of using filled drug prescriptions as a proxy for estimating disease level innovative. However, they also concluded that because diagnosis and prescription are closely related, the concern about a potential circularity in validation against self-reported diagnosis is real.

As we noted in the manuscript, illness and treatment are a series of steps:
1. The individual must recognize a medical need.
2. The individual must make the decision to seek medical care.
3. The individual must have access to medical care, including:
   a. Physical access (e.g., medical personnel within a reasonable distance),
   b. Physical access (e.g., access to a vehicle, especially as it relates to rural areas),
   c. Temporal access (e.g., accessible office hours, especially for minimum wage/single parent workers),
   d. Financial access (e.g., health insurance or sufficient cash),
   e. Social access (e.g., availability of medical facilities who serve minority populations),
   f. Social access (e.g., medical personnel who speak the patient’s language),
4. The individual’s condition must be diagnosed by medical personnel.
5. The individual’s treatment must include Rx (as opposed to strictly diet, exercise, etc.).
6. The individual must have access to Rx, including:
   a. Physical access (e.g., pharmacy within a reasonable distance),
   b. Physical access (e.g., access to a vehicle, especially as it relates to rural areas),
   c. Temporal access (e.g., accessible pharmacy hours, especially for minimum wage/single parent workers),
   d. Financial access (e.g., health insurance or sufficient cash),
   e. Social access (e.g., a pharmacy that serves minority populations),
7. The individual must follow the drug treatment regiment, including:
   a. The individual picks up the drug,
   b. The individual takes the drug on the proper schedule,
   c. The individual does not split the drug in order to double the doses,
   d. The individual does not share the drug with family members.

Additionally, the individual must perform in the following manner in order to be included in the BRFSS sample (our referent data set).
8. The household must have a land line (many have only cell phones)
9. The individual must be randomly chosen to participate in the BRFSS interview,
10. The individual must agree to participate in the BRFSS interview (could refuse, not be available, etc.),
11. The individual responds accurately to health status questions (i.e., no faulty memory, avoidance of questions, etc.)

We are aware that measuring Rx-filled rates and using it as a proxy measure of the prevalence of specific chronic illnesses is a crude methodology, fraught with possible disconnects, as the list above suggests. Nonetheless, we have accomplished what we consider to be a validation exercise. Can the measures/baskets of drugs be refined? Yes. Can other national data sets be
used for comparison? Perhaps, if funding and geographic identifiers permit. We do not feel that we have overstated the accuracy of this methodology, only that it is a promising methodology that enables policy makers and providers to visualize below the state level, and therefore the methodology should be further studied and refined.

To address the circularity, we realize that diagnosis and prescriptions are closely related; in fact, that is our argument. Small area estimates of diagnoses cannot be made for rural areas. We use prescription fill rates, which are available for small areas as a substitute so that some estimate of the spatial variation is available.

Though circular validity is potentially problematic, it is either circular validity that is the problem or it is the actual validity of using prescription drug data as a proxy. That is, either we can validate it with diagnoses data or it is an unacceptable proxy. We argue, and our data support, that it is a valid proxy and this is the initial test of that relationship.

2. Therefore, they feel for the paper to be considered further by PHM, this issue must be addressed - through validation against a "gold standard" at the national level (measured prevalence in NHANES, possibly by urban and rural place of residence and insurance status), and the paper should include a thorough discussion of the issue.

We have already presented an extensive test of correlation with a national “gold standard” in the form of the BRFSS. We are open to correlations test with other national health data sets in future funded research.

We are unable to use the NHANES, as suggested, as a comparison measure for the following reasons:

1. The health-related subsets of NHANES data do not ask appropriate questions with the exception of diabetes. The diabetes question is asked in the following manner: “Other than during pregnancy, have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?” This question is an appropriate indicator of prevalence of the disease. Unfortunately, there are no comparable questions for heart disease and stroke in any of the health-related subsets of data.

2. We cannot test the correlation of Rx data to diabetes NHANES data because there is no geographic identifying variable with which to merge the two datasets together. The method used in this manuscript merges the two datasets by state, and runs the correlation of Rxs filled at the state-level with the prevalence of disease at the state level. This is not possible using NHANES data because there is no state identifier attached to respondents in the publically available NHANES data set.

3. The only way this information can be accessed using the NHANES data is by gaining access to the Research Data Center. This is too time consuming and costly for us to perform considering the lack of robust comparison measures available. The following is taken from the NHANES website:

"Can I identify what region of the country or what state or county a survey participant resides within?"
3. We would be grateful if you could address this in a revised manuscript and provide a cover letter giving a point-by-point response to the reviewer concerns.

We have addressed the reviewer’s comments below (we changed the order of the reviews since #1 and #3 raise similar concerns).

Reviewer’s report
Title: Correlating pharmaceutical data with a national health survey as a proxy for estimating rural population health
Version: 1 Date: 3 November 2009
Reviewer: 1

Reviewer’s report:
Major Compulsory Revisions:
1. I am concerned that your reliance on IMS’ NDTI data leads to a high degree of misclassification of both medications and disease categories. You measure the number of "heart disease" prescriptions filled in a year in a state and compare this with individuals who report being told that they have "coronary heart disease".

We agree that there is some mismatch of drug to disease which will be addressed in future funded research. However we intentionally cast a wide net in terms of drugs that could/would/were used to treat specific chronic illnesses.

Your approach is complicated by the fact that, while some of the prescription medications you are counting are used for the treatment of individuals with coronary heart disease, all of them are used for the treatment of other, more prevalent conditions. Diabetes is relatively straightforward because diabetic medications are used, for the most part, only by diabetics. For stroke and heart disease, however this is not the case. For instance, the most common medications used for the secondary prevention of events in individuals with CHD (ASA, beta blockers, ACE inhibitors, and statins) are also the most common medications used for primary prevention of CHD in patients who do not have CHD. Individuals receiving them would not have been told by their health provider that they had CHD and likely would have been reassured, specifically, that they should take these medications to avoid developing CHD. This includes the vast number of individuals who are hypertensive (and therefore placed on beta-blockers, ace inhibitors, angiotensin receptor blockers, and calcium channel blockers) or hyperlipidemic (and therefore placed on lipid-lowering agents such as statins).

We agree that this is a measurement issue that needs to be addressed in future refinement of this methodology. Indeed, many drugs are used to prevent or slow the onset of chronic illness. One future solution might be to link Rx-filled data with a survey of physician’s prescribing practices (e.g., of your patients for whom you prescribe heart disease medications, what percentage are preventative versus treatment for existing disease?). To our knowledge no such data set currently exists.

More specifically we addressed the following issues when framing this methodology:
1. We worked with IMS Health to develop the basket of drugs. They are the leading firm in the collection, analysis and dissemination of prescription drug data in the U.S. They are aware of prescribing practices and trends in prescription drugs. Because the match between drug and illness is as much knowledge of what is occurring in the industry as it is the current medical guidelines, at the onset of the project, and very specifically, we chose IMS Health to select drug classes. (IMS: Based on the objectives for the research and the specific disease states it planned to address, IMS recommended specific categories of drugs, defined by IMS's Universal System of Classification-USC, to include in the data extracted and provided)

2. Prescribing practices change over time, as treatments are improved and new drugs are introduced. More specifically, the reviewers do not use past tense to refer to our 1999-2003 dataset. Our basket of drugs was the best practices for that time period.

3. Treatment practices are not nationally uniform, as clearly demonstrated in the Dartmouth Atlas of Health Care. Treatment practices change slowly over time, as measured by geography, the diffusion of best practices and the age of the practitioner. This slow change suggests that a wider, not narrower, net be cast when considering drugs for specific treatments.

4. When flows of individual drugs are examined, altering one or two in the calculations would have no effect on the overall conclusion of this manuscript which is; here is a new and potentially valuable methodology to measure population health metrics, warts and all.

Furthermore you count alpha blockers, which are most commonly prescribed at low dose for benign prostatic hypertrophy and only rarely used for hypertension anymore.

The use of “anymore” strongly suggests that the reviewer is referencing current practices, as opposed to prescribing practices from 1999 – 2003, which is the time period for the data set.

This leads to a further concern, namely that you may not have, as you state on page 9, "paired drug classes with appropriate BRFSS questions". For instance, you estimate correlations between heart disease prescriptions (which include a large number of lipid-lowering medications) and individuals with high blood pressure (a condition not treated with lipid-lowering).

We agree with the reviewer’s observation about the paring. In fact the correlation was higher between high blood pressure and heart disease prescriptions filled (0.733 in 2003) than the correlation between coronary heart disease and heart disease prescriptions filled (0.613 in 2003). At this point in our research we cannot account for this finding. However we felt that, in the spirit of full disclosure of research results, we would report all combinations. We plan to explore this finding further in future funded research.

2. NDTI offers a confusingly broad list of medications for cerebrovascular disease, including: oral anticoagulants (i.e. warfarin) that are actually used for atrial fibrillation and only rarely to treat individuals who have suffered a stroke, oral-antiplatelets (aspirin, clopidogrel and dipyridamole) that are used widely for a many different conditions including coronary heart disease as well as stroke, Vitamin K which is actually used to reverse anticoagulation, as well as dementia medications, anti-Parkinsonians, and anti-epileptics. If your methods work, this list will still lead you to dramatically overestimate the prevalence of those having suffered a cerebrovascular accident (stroke). Furthermore, it is unclear why you would compare this population with individuals who have hypertension (Table 1).

While some individuals who have strokes are hypertensive, many are not. Your attempt to estimate stroke is complicated by the widely varying treatments for ischemic, embolic, and hemorrhagic stroke. You state on page 18 that "stroke medications were used to treat a fraction
of hypertension cases.” Why is this? - none of the medications you count for cerebrovascular disease are anti-hypertensives.

See points made above.
(IMS: As with many prevalent disease states, treatment for hypertension in the U.S. is complex and evolving. Recently available data at the anonymized patient level show evolving treatment algorithms, use of multiple drugs concomitantly, use of combination drugs, and switching of patients among therapies in order to reach treatment goals)

3. According to your methods section, IMS data misses 28% of prescriptions but a proprietary method is used to estimate the remaining missing data. It would help to understand how this method does (or does not) account for the 23% of prescriptions missing by mail order and how this is likely to bias your results.

“Retail pharmacies account for 67% of total national prescription sales.” (p. 8) with 23% via mail, 8% in clinics, etc. The reviewer may have misinterpreted this to mean that the remaining 33% was missing from the total. That is not correct. IMS Health is aware of the location and those Rx are assigned to their proper geography. We agree that we would prefer knowing their methodology, but given its’ proprietary nature, it cannot be shared.
(IMS: IMS’s prescription data base, known as Xponent at the physician level and National Prescription Audit at the national level, includes data from Retail pharmacies, Mail Order pharmacies, and Long-Term Care facilities, the three channels for which prescription data exist. The data bases result from collection of very large samples from outlets in these three channels, and projection to physician and national totals for outlets not included in the extensive samples. The data collection and projection methodologies are considered to be among the best employed for the creation of any commercial data base. The projection methodology is patented.)

4. Given the widespread use of antihypertensives and lipid-lowering medications, your estimates of filled prescriptions at the county level appear much too low. If these are accurate, it would help to contrast them with other available estimates and discuss why they differ so dramatically.

On the basis of what data sets does the reviewer feel that our estimates are antihypertensives and lipid-lowering medications? Is the reviewer using current knowledge or data from 1999-2003?

Minor Essential Revisions:
1. You refer to figure 1b on page 12 but have no figure labeled as such.

That should have referred to Figure 2. We have corrected that in the manuscript.

2. Many physicians provide 3 months of medications at a time, and many individuals fill their prescriptions once every 3 months, but you assume that each prescription is for 30 days (since you divide by 12 months). Other individuals fill their prescriptions for only part of each year. How do you account for this?
The 3-month order was not widely available in 1999-2003. We cannot fully account for those who only fill their prescription for only part of the year beyond our rolling 12-month calculation. (IMS: As an example, in 2003, prescriptions in the Retail channel, which were overwhelmingly 30-day supplies at the time, accounted for 86% of all prescriptions for the Angiotensin Receptor Blocker drug class. Mail Order prescriptions, which are typically for 90 days, only accounted for 10% of ARB prescriptions, and Long-Term Care accounted for the remaining 4% of ARB prescriptions.)

3. Your results and discussions sections need to be separated.

*We have made revisions to the text.*

Discretionary Revisions:
1. In their background, the authors mention NHIS and BRFSS as sources of nationwide prevalence data but do not mention other well-known sources: NHANES, CMS/Medicare data, and National Hospital Discharge Survey.

*We did not consider NHANES since it does not have measures for heart disease and stroke/hypertension. CMS/Medicare is not representatives of the entire population. The same for hospital discharge data since the majority of those with chronic diseases are not admitted to a hospital.*

2. I am concerned that differences between your maps and BRFSS maps are not "suggestive", as you state on page 19, as much as a reflection of problems with your methods. I fear that you may be comparing very different patient populations.

*That is an issue anytime a new methodology is used. Further refinement of the methodology will determine if the differences are real or an artifact of the methodology.*

Level of interest: An article of limited interest

Quality of written English: Needs some language corrections before being published

Reviewer's report
Title: Correlating pharmaceutical data with a national health survey as a proxy for estimating rural population health
Version: 1 Date: 13 February 2010
Reviewer: 3
Reviewer's report:
Major Compulsory Revisions:

1. The results section is a mixture of background, methods, results, and discussion and should be revisited so the results section is limited to presentation of results.

*We have revised the manuscript.*

2. As noted by another reviewer, the medicines for heart disease and cerebrovascular disease in particular are also used for other indications. This should be addressed and further
emphasized at the very least as a limitation. Was the selection of the medicines subject to a medical or pharmacist review?

We made a decision to avoid practitioner bias and rely upon choices made by business analysts in the field of Rx flows and treatments (IMS Health). They are in a position to see national practices and are not restricted to their individual clinic or training. (IMS: IMS based its recommendations of drug classes on those primarily used to treat the disease states being researched (i.e., not extensively used in other disease states), as well as on those which would provide the majority of prescriptions for these disease states. We referenced another IMS data base-National Disease and Therapeutic Index- which tracks physicians’ use of drugs by indications, desired action, etc. to help form these recommendations.)

3. I question the value of including all of the figures.

In this case, a finding of no significance (no positive spatial autocorrelation) is significant. Many in the rural health field believe that physical, logistical and financial barriers prevent adequate care in rural areas, which these maps clearly show is not the case nationally for prescription drugs. Furthermore, the maps are valuable to those who wish to inspect sub-state spatial patterns, which was not a focus of this manuscript.

4. What additional validation studies do the authors recommend?

We have added this to the manuscript, specifically on page 22, Future Directions.

Minor Essential Revisions:

1. Clarify if this was exploratory or if there were a priori levels of correlation that the authors were seeking.

This study was entirely exploratory. There were no a priori expectations.

2. Indicate what percentage of adults receive 3-month supplies of chronic medications then do a sensitivity analysis on its effect on results.

Prescriptions are reported on a monthly basis and were not broken out by length of time since 3-month supplies were not common in 1999-2003. (IMS: The prescription data provided to support the research included Retail and Mail Order prescriptions, but did not break out prescriptions by these channels.)

3. The title indicates the study is about rural population health, yet it seems that the study encompasses all U.S. settings. The title should be adjusted accordingly.

The study is intended to provide a chronic illness prevalence measure at the county level, which is more useful in rural settings. The paper compares urban to rural areas.

Small-area estimates can be made in urban settings where survey sample sizes are large enough. The emphasis on rural is because these small-area estimates will allow for a better understanding of variation in rural areas.
4. There are a few misspellings in the manuscript.

*We have edited the manuscript carefully.*

5. A better description of the IMS Health, Inc., datasets should be included.

*We believe that we have adequately described the data sets in our three-page explanation (p. 6-8), as well as our reference to a fuller description (footnote #28).*

**Discretionary Revisions:**

1. What is the cost of using the IMS Health, Inc., datasets? This may be of interest should others wish to use this approach.

*This data set was a custom data run. Interested researchers are urged to contact IMS Health directly for a data quote.*

(IMS: IMS provides data sets of both a custom and syndicated nature, to support the activities of pharmaceutical companies, financial institutions, academic researchers and other categories of potential users.)

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

---

**Reviewer's report**

**Title:** Correlating pharmaceutical data with a national health survey as a proxy for estimating rural population health

**Version:** 1  **Date:** 26 January 2010

**Reviewer:** 2

**Reviewer's report:**

There are no major compulsory revisions

**Minor essential revisions**

1. The source of the population data used as denominators needs to be cited (i.e., the official US Census Bureau post-censal population estimates (See # 2 for an online entry point to these data).

*We have revised the manuscript.*

2. A brief discussion of the method used by the US Census Bureau to develop county-level population estimates by age needs to be included. The brief description can be extracted from material found at [http://www.census.gov/popest/topics/methodology/](http://www.census.gov/popest/topics/methodology/)

*We have revised the manuscript on page 22.*

3. In considering future directions (i.e., sub-county prevalence rates), the authors should state that the US Census Bureau estimates cannot be directly used for this purpose, but that in conjunction with synthetic estimation and other techniques that the "cohort change ratio"
technique can be used for this purpose (See, Smith, Tayman, and Swanson, 2001; Swanson, Schlottmann, and Schmidt, 2009).

References

We have revised the manuscript and added these citations on page 22.

Level of interest: An exceptional article

Quality of written English: Acceptable