Author's response to reviews

Title: The Health Disparities Cancer Collaborative: a case study of practice registry measurement in a quality improvement collaborative

Authors:

    David A Haggstrom (dahaggst@iupui.edu)
    Steven B Clauser (clausers@mail.nih.gov)
    Stephen H Taplin (taplins@mail.nih.gov)

Version: 2 Date: 5 February 2010

Author's response to reviews: see over
Dear Dr. Mittman:

Please accept this revision of our manuscript, entitled: “The Health Disparities Cancer Collaborative: a case study of practice registry measurement in a quality improvement collaborative.” Attached is an itemized, point-by-point response to the comments of the reviewers. Changes made to the manuscript are indicated in *italics* throughout. We sincerely appreciate the excellent, thoughtful reviews that we received and believe this feedback has made our manuscript better. If you need any further clarification or materials, please do not hesitate to contact me at (317) 988-2067. You can also email me at dahaggst@iupui.edu.

Best Regards,

David A. Haggstrom, MD, MAS
Core Investigator, VA HSR&D Center on Implementing Evidence-Based Practice
Assistant Professor, Indiana University, Division of General Internal Medicine
Scientist, Regenstrief Institute, Center for Health Services & Outcomes Research
Reviewer #1:

1. More information about the completeness of the registry would be quite useful in understanding its utility as a measurement tool. It sounds like each center had to enter into the registry all patients who had been seen in a three-year time period before the HDCC; do the authors know how long it took them to do this, and do they have any validation of the completeness of this data entry? Were these patients entered in chronological order, or was there some sorting/prioritizing of those with significant medical problems, abnormal screening tests, or other characteristic? If the calculation of screening rates was based on who had happened to be entered in the registry by a particular point in time, the pattern and completeness of this data entry could matter a lot.

The reviewer raises several good points, and in fact, the inability to answer these questions is the basis for the cautionary tone of our discussion. The question is pertinent to the interpretation of our data as well as the implementation of measurement in practice. We do not know how long it took each center to enter patients into the registry, and data entry was variable. We now better describe the range of data entry processes in the methods: “data entry varied from the wholesale transfer of demographic information from billing data queried for age-appropriate groups, to hand entry (p. 8)”.

Implicit in this question and explicit in our discussion is the fact that the data entry task may be a large one: in the discussion we state “the experience of the HDCC suggests that the data entry burden for large screening populations poses significant challenges for primary care practices”. We have now added that “formal assessment of the burden of data entry and tracking activities upon health center personnel would inform estimates of the cost of other collaboratives targeting large populations (p. 20)”.

We also do not have validation of the completeness of this data entry. Furthermore - outside of age eligibility for screening (described in performance measures section) and having had a visit within the previous three years - there was no other uniform sorting/prioritizing that influenced how patients were entered. In the discussion, we state that “the validity of the aggregate findings regarding cancer screening are uncertain. Heterogeneous methods of practice registry data collection across a heterogeneous group of health centers (different sizes and approaches to care) limit the confidence with which the pooled data can be interpreted and compared to outside organizations (p. 22).”

a. Given that much of the data required for entry into the registry depended on reports from outside institutions (e.g. mammogram reports), what was the workflow such that these data were entered into the registry on an ongoing basis? And again, any validation of completeness?

The dependence of community health centers upon data from outside institutions is an important insight, and one that the paper takes head-on. First of all, we note that “Both the small number of events reported, and centers who reported them, commonly made it infeasible to test for statistically significant changes in follow-up or treatment, even over the entire collaborative (p. 14)”, and that one of the primary explanations may
have been “health centers did not have routine access to the medical information necessary to report the measures because the care occurs outside their practice (p. 14)”. A subsequent section of the discussion is then dedicated to this topic: “Why health centers may not have access to the data necessary to report the measures (p. 16)”. In response to the reviewer’s comment, we now also observe that “there was no uniform workflow for data from institutions outside the HDCC (p. 16)”. Again, we did not perform additional validation of the data; and consequently, the abstract conclusion state that “more definitive evaluation would require validation of the registries.” We have also added the following regarding validation: “Quality improvement efforts do not routinely perform data validation, although strategic data quality checks would be worthwhile (p. 23).”

2. The delineation of the three patterns of change in numbers and proportions of patients screened is a useful framework, and provides a good model for reporting of results in future studies of population screening efforts. However, it would be interesting to see if there were any center-specific patterns in these changes – i.e. if one center had a particular pattern of change for breast cancer screening, was it the same for the other 2 types of screening as well? Do patterns of change in all 3 types of screening tend to track together at each individual center? What do these patterns mean about particular QI strengths/weaknesses/strategies at the site level?

We previously showed the patterns of change at the center level in the appendix (Additional file 2; now Table 5 per Comment 33). The reviewer raises a very interesting point, and it is worth noting that patterns of change do tend to track together at each individual center to some extent.

We have now described these patterns…

“At the individual health center level, patterns of change tended to track together across the three types of screening. At two centers, the second pattern of change (Figure 2) occurred across breast, cervical, and colorectal cancer screening, and at another center, across breast and cervical cancer screening. At two centers, the third pattern of change (Figure 3) occurred across both breast and cervical cancer screening. (p. 13).”

…and provided an interpretation in the discussion:

“The observation that unique patterns of changed tracked across difference cancer screening tests at the same center further suggests that explanations related to data collection and entry most likely drive these patterns. (p. 19).” We also proceed to discuss why centers may experience early issues with data entry (response to comment 28), as well as what other biases may be introduced by the data collection process (response to comment 23).

Minor Essential Revisions
3. Can the authors tell us more about these 16 health centers? Are they typical of health centers around the country? How were they selected to participate in the HDCC?
We have now provided a table (Table 1) containing descriptive statistics of (mean, range of) patients eligible for each cancer screening test, the number of doctors and nurses, the number of months reporting registry data, as well as the geographic location. We also now describe how centers were selected to participate in the HDCC: “The 16 centers were selected through an active process that involved telephone interviews with health center leaders to assess their enthusiasm and willingness to commit the resources necessary for success (p. 6).”

4. Given that the aim is in part a real-world description of how practice registries work, some more information about time/staff needed for data entry and registry maintenance (and how these resources were funded) would be useful. How many centers used the HDCC-provided registry software, and what did those that did something else do? How many of the centers had an electronic medical record (EMR) or other electronic patient accounts system – and did these automatically flow into the HDCC patient registry?

We have now clarified our previous statement in the methods about the practice registry: “All health centers participating in the HDCC used the practice registry data software provided by the HDCC; nationwide, HRSA community health centers were encouraged, but not mandated to use the software (p. 7-8).”

We do not know how many of the centers had an EMR. “Although some centers performed automated data transfers from billing systems to registries, this process required advanced data management capabilities that were not always available (p. 20).”

5. Do Figures 1, 2, and 3 show data for just breast cancer screening, or for all 3 screening types? The text on p.11 where the figures are referred to made me expect that the figures would show the patterns for all 3 types.

To page 11, we have now added the parenthetical comment that the figures use “representative breast cancer screening examples from an individual health center”. These patterns of change could occur across screening types other than breast cancer screening, but the figure examples are chosen so as to be representative of different patterns of practice change.

Discretionary Revisions / Comments

6. Page 6 could benefit from a little more detail about what these “local teams” looked like and how much staff time/effort/$ was allocated to these efforts.

We have now described the composition of the local teams: “The local teams consisted of employees with multiple backgrounds and roles, including providers (physicians, physician assistants, and nurse practitioners), nurses, appointment staff, information systems, and laboratory personnel. The effort and staff time allocated averaged 4 FTE/team with an aggregate of 950 hours/team (p. 6).”

7. The concept described in analysis #3 (page 9-10) – i.e. looking at both the n and the % of pts screened, as well as the size of the target population – is practical and useful. Clearly, when investigating the drivers of quality change (or lack of change) at individual
institutions, one needs to keep all of these items in mind. There may be merits for particular situations/institutions to choosing one indicator over another as a benchmark to publicize, and these choices should be made consciously and carefully.

We agree and appreciate the reviewer’s comment about the practicality and usefulness of these types of considerations.

8. One aspect of the data presented in Table 2, in addition to the main point discussed in the text (i.e. detectable change), is the disturbingly low rate of timely treatment after cancer diagnoses. I understand this chart to say that for the 31 women diagnosed with breast cancer, only 2 received initial treatment within 90 days of diagnosis; similarly slow follow-up seems the case for CIN2/3 and colon polyps/cancer. Can the authors comment on the lack of timely treatment?

These rates are very low. In the discussion, we have commented further on the apparent lack of timely treatment among reporting health centers, and given an honest accounting of what most likely explains these findings: “the extremely low rate of timely treatment after cancer diagnoses among reporting health centers (3-24%) very likely represented the lack of a systematic way to collect feedback from oncology practices rather than quality gaps; data across practices is very difficult to locate outside the context of integrated data and delivery systems. Health centers appeared to report what little information was available regarding follow-up and treatment and shift their focus to cancer screening. In the subsequent HDCC regional collaborative, substantial emphasis was placed upon building communities of practice to help address the lack of coordination between primary care and subspecialty practices (p. 16-17).”

9. On pages 13-15, the authors’ third explanation for lack of data on the more ‘distal’ events in the cancer screening process seems the most plausible. I would assume that mammography and colonoscopy do not occur on-site at these health centers, and it may often be difficult for centers to receive initial test results – let alone information on the follow-up of abnormal tests. I might have expected more data available in the registry for Pap smears and colposcopies, as I would imagine that many of these are performed on-site at the centers. Do the authors have information on how many of the centers offered these services on-site, and if data quality/completeness varied accordingly?

We agree that the third explanation is the most likely, and have now included this reasonable opinion in the discussion: “we suspect that the lack of access to data outside the primary care practices contributes most to the small number of abnormal screening results and cancer diagnoses reported (p. 16)”.

We also agree that it is a reasonable to expect that health centers were more likely to perform Pap tests on-site and have now made the following observation: “Of the screening follow-up steps reviewed, the highest percentage of health centers reported timely notification of Pap test results (62.5%), most likely because these services were performed onsite at the health centers (p. 15).” We do not have empiric information about how many of the centers offered these services on-site.
10. I would agree that for comprehensive accountability (p.15) we need health-system change and integration across sites – any specific thoughts as to how to make this happen? Are ‘patient-centered homes’ with resources available for coordination the answer?

Our previous draft suggested the policy response of accountable care organizations, i.e., “reward the team of providers responsible for the care of patients with complex medical conditions, including cancer (p. 17)” Given that community health centers are primary care practices, patient-centered medical homes are also a natural policy response to the issues of coordination and data tracking highlighted in this paper, thus, we have added the following statement: “Policymakers may also want to consider our findings as reinforcing evidence of the potential value of patient-centered medical homes, if they make additional resources available for coordinating care with other providers and using data systems to track referrals and results (p. 17).”

11. If the authors have any information about particularly successful health centers in the HDCC, that could be interesting to present, even in an anecdotal form.

We think this would be interesting as well, but are concerned that adding these anecdotes may confuse the already complex organization of this paper.

12. I would also agree with the authors’ comments on pages 16-17 re: dangers of an unreliable denominator. Until facilities are equipped with comprehensive EMRs, the burden of maintaining a patient registry to track large-population tasks like cancer screening seems overwhelming and likely to fail (be inaccurate/incomplete) due to limited resources. The HDCC is probably the best-case scenario, given that these were probably particularly motivated centers with special attention from national organizations.

We agree with the first half of the comment, as well as the estimation that “the HDCC likely represented a best-case scenario of particularly motivated health centers with special attention from national organizations (p. 19)”.

13. I would strongly encourage a table that summarizes basic data about all 16 health centers

We have now provided Table 1 containing descriptive statistics of (mean, range of) patients eligible for each cancer screening test, the number of doctors and nurses, the number of months reporting registry data, as well as the geographic location.

14. I would also encourage more description of what % of clinics used the HDCC registry software, and what kind of training was provided during the Collaborative in how to set up, populate, and sustain useful registries.

All health centers used the HDCC registry software. We have added the following information about training: “Training in the software was provided by HRSA
at a national level, as an adjunct to collaborative learning sessions, and at the regional and local level by the Information System Specialist (ISS). The training typically consisted of 4-8 hour interactive sessions in which participants would have a “live” experience on laptops (p. 8).”

15. Also a bit more discussion of what appears to be anomalous registry results (Fig 2) would be helpful.

We’re uncertain what particular anomalous results are being referred to, although in the previous manuscript, we mislabeled Figure 2 as Figure 3 (Comment 27). We have also now commented more upon the Figure 2 results (Comment 31).

Review: Health Disparities Cancer Collaborative: a case study of practice registry measurement in a quality improvement collaborative.

16. This paper examines the quality of data in cancer registries of 16 community health centers taking part in a quality improvement collaborative (the Health Disparities Cancer Collaborative). As the authors mention in the conclusion, they have provided an “unvarnished” experience with the Health Disparities Collaborative. With that in mind I think they could provide some more details regarding the process by which health centers were directed/encouraged to create registries, e.g. what training did they receive, what percent used the HDCC software, what kind of staff person was in charge of the registry at the health centers, and where were data posted each month.

The request for more details about the registries was a recurrent theme in the reviews. In response to previous comments, we have addressed who used the registries (Comment 4), and how they were trained (Comment 14). We now provide further historical description of how HRSA directed registry development among health centers:

“In 2000, HRSA supported the development and deployment of an electronic registry software. Over the next five years, HRSA continued to support numerous iterations of the registry software to address both the increasing scope of the collaboratives (such as cancer screening) and the needs of clinicians and other frontline-staff users. Informing this process was an advisory group of health center clinicians and technical experts which provided insight and guidance about critical registry functionalities and the needs of measurement to effectively support practice management (p. 8).”

We also have added further information about where data was posted each month:

“The data was posted on a secure data repository to be shared with HDCC facilitators and benchmarked against other health centers. A data manager from the medical records department at each center who had training in use of the registry uploaded the data (p. 8).”

17. The manuscript brings an important “inside” look at the process of disease registries which are a common feature of quality improvement collaboratives. The design and methods are solid, and the discussion is thorough with useful lessons for quality improvement methods and evaluation. It is very well written.
Thank you. We take special pride in the quality of the writing and take this as a high compliment.

**Background:**
18. P 5: “half of physician organizations” -- consider “physician office practices” instead (I thought of physicians unions first, but I don’t think that was meant).

This language may be awkward, but it is drawn directly from the reference so should probably be preserved:


19. P5: Good point to examine completeness of data across health centers over time: in some collaboratives registries start out very small (e.g. 10-20 patients) and as they are added to, the results can vary greatly (which seems to be the case here too).

It is validating to hear that this type of pattern holds across other collaborative experiences.

**Methods:**
20. What percent used the HDCC software for the registry? Were any analyses done looking at completeness of data and improvements, comparing the centers that used the HDCC software to those that did not.

All of the health centers used the HDCC software (p. 7), thus, there was no variation to perform a comparison between centers using the software and those that did not.

**Results:**
21. I would find an introductory table helpful that describes the 16 CHCs participating, e.g. basic statistics on number of patients (e.g. mean, range, standard deviation), number of doctors, number of nurses, geographic location, whether have an EHR, used the HDCC software or not, number of months reporting registry data, etc.

We have now provided Table 1 containing descriptive statistics of (mean, range of) patients eligible for each cancer screening test, the number of doctors and nurses, the number of months reporting registry data, as well as the geographic location. We don’t currently have data about number of patients seen overall by the health centers. We now clarify that we do not know whether centers had an EHR, but that all participating centers used the HDCC software (Comment 4).

22. P7. How many of the 16 centers used the HDCC registry software? For those that didn’t what systems were used?

See previous comments 4, 14, and 16.
23. *P8. the process of entering patients into the registry has substantial potential for bias, i.e. those more active patients (seen in one of the months of the collaborative) would be more likely to be entered into the registry than patients who had not been seen in some time. Also the more active patients would be more likely to have the screenings and follow-ups, assuming that those were issues covered in the collaborative sessions.*

This is a very important issue. We now describe this selection bias into the discussion in a manner similar to that suggested by the reviewer: “The process of entering patient data into the registry also has substantial potential for bias, i.e. those more active patients (seen in one of the months of the collaborative) would be more likely to be entered into the registry than patients who had not been seen in some time. Also the more active patients would be more likely to have screening and follow-ups, given that those were issues covered in the collaborative sessions (p. 21).”

A closely related issue is the impact different definitions of the denominator population eligible for screening have upon screening performance. This related topic has been reported in the literature since our initial submission, and because of its clear applicability to our findings, we now describe this issue in some detail.

“There is a reasonable expectation that the relatively inclusive sampling approach to the HDCC’s eligible denominator population (seen once in the past 3 years) underestimates the screening performance, compared to less inclusive sampling approaches to the eligible screening population (for example, if patients were included only if they had been seen in the past year). Practically speaking, even though the eligible denominator population was standardized and health centers were encouraged to enter that denominator at the beginning of the collaborative, the burden of data entry was considerable, and not all health centers likely could establish the full eligible population by day one of measurement. Thus, centers may have initially been including eligible individuals seen in only the past few months or year. With a less inclusive sampling approach of this type, these centers likely overestimated screening performance. Yet because assessment in the collaborative was primarily done for internal quality improvement, not external reporting purposes, a more inclusive definition of the eligible population was desirable because it can afford centers the opportunity to identify patient populations that might benefit from more intensive outreach (p. 21-22).”

24. *P10, last line – “25%” should be “24%”, according to Table 2.*

We have changed the percentage in the text to be consistent with the correct percentage found in the table.

25. *P11, For Figures 1, 2, and 3, it would be helpful to know how many centers were involved. Since all three figures are just for breast cancer, the sum of the numbers of CHCs represented in Fig 1, 2, and 3 would equal to 16—if I’ve understood the numbers correctly.*

For figures 1, 2, and 3, only an individual health center is involved; the title of each figure has been edited accordingly. These figures are intended to serve as
representative of how the data may appear at the health center level, and the text of the manuscript has been edited to make this point (p. 11), as outlined in comment 5.

**Discussion**

26. Important finding that selection bias threatens the validity of general inferences drawn from the overall collaborative. This points out one of the challenges of evaluating collaboratives – i.e. having good quality data of processes and outcomes, and secondly having comparison clinics. As is pointed out later, there were no comparison clinics, so it is not clear that even the improvements in proportion of patients screened can be attributed to the collaborative, and not to other trends going on in community health care centers at the time. The discussion is good in that it draws out the implications, and limitations of these data, e.g. health organizations may be hesitant to report data unless they are favorable. Also very interesting and important point about the importance of collecting data not just at individual community health care centers, because of care and services that are provided at specialty referral practices. This points to the large gap in the US health care system in which data for individual patients who receive care from different health care organizations usually are poorly linked.

This summarizes some of the key points made in the discussion, and we believe, highlights the value this paper would contribute to the medical literature.

27. **P16.** In the third line from the bottom, I think the authors mean Figure 2 (not Figure 3) where there is a late, rapid increase in eligible population.

We apologize for this oversight and have made the appropriate correction.

28. **P17:** I agree that it is challenging to define the denominator population, and that many clinics do not have disease registries. The authors have done a nice job of raising some of the inconsistencies – however one point they don’t raise is that this is probably a new process for many of the clinics. They may have gotten a brief “lesson” about disease registries during one of the three collaborative sessions, but it may have been insufficient for them to be proficient and accurate. They are likely to make mistakes at first, and thus I think there may be a lot of “noise” in the first year of data. I think Figure 2 exemplifies this with the dramatic shifts in every category – number screened, eligible population, and proportion screened. (see comments below about Figure 2).

The reviewer raises another good point, and we have now incorporated their thoughts into the discussion: “Using registries to track screening is a new organizational process for many practices. These centers received training, but training does not replace actual practice experience in allowing organizations to become proficient. Practices are likely to encounter problems at first, and thus, there may be considerable imprecision in the first year of data (p. 19).”

29. **P17:** “...data entry burden for large screening populations...” It would be helpful to describe in the “Collaborative Intervention” section how much training was given in disease registries, and something about what the HDCC software was like – e.g. based
upon an existing software like Excel or completely homegrown? Had it been used in other settings?

Training described in Comment 14. The electronic registry was homegrown (developed and deployed by HRSA, p. 8), and had not been previously used in other settings.

Figures:
30. It would be helpful to add in the legend/title the number of centers to which each figure applies. If I understand Figures 1, 2, and 3, in total, they sum to the 16 centers.

Same comment (and response) as comment 25.

31. Figure 2: there is a striking anomaly in the May to October period – large drop in number eligible and number screened. What might explain this? Could this be due to one large center (with large number of patients) where the procedures for keeping the registry changed for some reason?

During this period, there is a large (rise and) drop in the number eligible, but not screened. We now highlight this anomaly in our discussion: “the sharp rise and drop in the number eligible midway through the HDCC in Figure 2 likely represents a mid-course correction in how the eligible, denominator population was ascertained (p. 18-19).”

Author Contributions:
32. In the second sentence one author, ST, is listed twice – I’m assuming this is a typo.

Observation is correct: we have changed the sentence to list DH, SC, and ST.

Additional File 2:
33. I would consider including this in the main paper, or at least the screening mammography data.

If the editors agree, we will include Additional file 2 as Table 5 in the main paper.

Again, this shows the “messiness” of use of registries and new data monitoring procedures for community health centers. For example the drop of 497 patients eligible for screening mammography. This was probably a center where they misunderstood the processes and criterion that were outlined in the collaborative, and maybe through one of the phone conferences or in a one-on-one with collaborative faculty learned that their methods were different from the other programs – and thus this adjustment was made. I think it also would suggest, and the authors might consider adding this to the discussion, that if collaboratives are thinking of using these kind of data as an evaluation tool of how well the collaborative is working, then more effort is needed to bring all the clinics up to speed. This has been advocated before in other articles, but it bears repeating.
Collaboratives often have a heterogeneous group of clinics – from different regions, different sizes, different approaches to care. Getting them all to use similar registry and reporting procedures can be quite challenging.

We want to be clear that the HDCC leaders did not propose to use these data as an evaluation tool without considering the importance of training or clinic heterogeneity. In fact, the opposite is true - the HDCC leaders allowed our research team to use this data to illustrate these measurement-related issues - in the spirit that other organizations could learn from “our unvarnished experience (p. 23)” and consider some of the lessons learned to better understand what may be the strengths and limitations from such data collection efforts.

We agree with, and have already highlighted, the importance of heterogeneity: “Heterogeneous methods of practice registry data collection across health centers limit the confidence with which the pooled data can be interpreted and compared to outside organizations (p. 22).” To further emphasize the amount of heterogeneity present, we have added the following description to the text: “a heterogeneous groups of health centers (different sizes and approaches to data entry) (p. 22).”

We have also added this follow-up comment to the discussion: “Before external audiences use this type of data as an evaluation tool of an overall collaborative’s performance, standardization in the training and experience with the registry is necessary, as well as critical thought about how to consider the various types of heterogeneity across organizations (p. 22).”