Article details: 2016-0104	
Title	A population based controlled before and after study of the impact of a provincial quality improvement program on primary health care in Ontario
Authors	Michael E. Green MD MPH, Stewart B. Harris MD MPH, Susan Webster-Bogaert MA, Han Han PhD, Jyoti Kotecha MPA MRSC, Alexander Kopp BA, Minnie M. Ho MHSc, Richard V. Birtwhistle MD MSc, Richard H. Glazier MD MPH
Reviewer 1	Dr. Cheryl Ann Levitt
Institution	Department of Family Medicine, McMaster University, Hamilton, Ont.
General comments (author response in bold)	This is generally well written and very nicely executed before and after study using comprehensive linked administrative datasets. From my read, the study was undertaken to determine if the QIIP LC (given three times to separate groups), using the IHI breakthrough series methodology, improved diabetes process of care, colorectal and cervical cancer screening and health care utilization (a proxy measure of access). There were some improvements in diabetes process of care and cancer screening but no improvements for health care utilization. The abstract is well written. The large number of physicians studied and patients included is impressive. A number of issues could be edited to improve the coherence and readability of the paper: 1. CQI and IHI should be spelt out in the abstract. 2. Use of the words "IHI breakthrough series" in the abstract is not explained in the paper. Could be on p 6 line 15., or left out 3. When did QIIP get established and when did this QIIP LC take place (x3)? P6. Line 13. 4. Methods: Can you describe what the LC was and how it differed from the control? It's not clear from the methods what the three LCs were and how they aimed to improve practice. 5. Need consistency and follow through with terms used. a. The use of the terms formally enrolled, rostered, panel, are not consistently used and not well explained for audiences not in Ontario. Also, the grammar can be improved. P7. Line 20 "Patients were assigned (rostered) to a physician if they were (agreed to be) formally enrolled as a patient with that physician." … rest of sentence needs clarifying. Not sure where "basket of services" fit in, or if you need to include this at all. b. You should explain that these rostered patients also agree to have their administrative data collected and shared for research purposes. There would be other patients that are not formally rostered? They would be excluded as their administrative data is not collected. c. Not sure that "panel" needs to be used at all. If so, perhaps you can explain what a "panel" is. Is it the list of patients rostered with a physician? Is 100 patients - a 100 rostered patients? 6. P7 line 40. "so all non-consenting eligible physicians made up were included in the control group." I don't think this is correct? I think they made up 5% of the control group (which would be helpful to explain in the results). What about the 10 refusing physicians, were they excluded completely or were they included in the controls as you did not know who they were? That whole paragraph can be made clearer. 7. Statistical analysis. P8. I think it would be useful to describe what p value was regarded as statistically significant and highlight those in the Tables. I think the statistical description is very brief and could be stronger. The unit of analysis was the physician or the practice? 8. 10 physicians were ineligible, means you were able to identify them but not the refusing physicians? Please explain. 9. P9. Line 36. What are the 3 index dates? The explanation of the three LCs in the methods might clarify this. 10. P11. Line 4, continuity of care … this is not in the table, but explained here. Need to clarify that it is not in the table. Perhaps a separate paragraph? 11. P11. Line 32. "It should be noted…impact" This sentence is not very clear and needs improving. 12. You must comment on the controls improving too and the many efforts taking place simultaneously in all of family medicine, not only FHTs, to improve diabetes and cancer. 13. P11 last paragraph line 54..will be easier to read if you include the comparison numbers from this study, otherwise I kept having to look at the Table 14. P12. Line 42. Need better numbers, add, include those in the table, not just the differences. 15. Limitations: at last, I got the picture of the problem with the LC participants. Some of this could be in the methods as it bothered me reading through the paper. 16. Not sure what a "program implementation rather than a trial of an intervention" means. Please clarify. I thought the intervention is the LC. Reference? 17. P14. Line 6. To mitigate… good sentence. Do you have a reference? 18. P14. Line 18. This repeating of the analysis later in time might also show how sustainable the intervention was…. But logically, with the ongoing changes in policy, leadership, direction, evidence, it might not be relevant later on. 19. P14 line 38. Use of these proxy outcomes. Have you a reference? This could be better explained. Logical that getting seen in a timely manner (same day) would result in less use of emergency room visits and better continuity of care. But if some of the problems people have go away over time and they don't need to be seen at all, perhaps these outcomes won't show the changes? Perhaps administrative data like this is a poor proxy. 20. P14. Line 50. "in our settings…" Not sure what you mean here. **Many of these comments were addressed in response to the editors comments for example explaining the program, eliminating excessive use of abbreviations. In addition:** **1. Section on assigning patients and terminology on enrollment was revised.** **2. The comment in 5b is incorrect. No patient consent is required to use this data and we have data on all patients whether formally enrolled or not.** **3. Language on inclusion criteria for the groups was cleaned up.** **4. Added a line in limitations expanding on the specific provincial initiatives in diabetes and cancer care.** **5. Clarified meaning of "setting" in the conclusion.** **6. Clarification added to methods on cut off for statistical significance. It is already stated in this section that the unit of analysis is the physician.** **7. Clarified statement on program implementation and risk of bias in the limitations section.**
Reviewer 2	Dr. Susan Baxter PhD
Institution	Faculty of Health Science, Simon Fraser University, Burnaby, BC
General comments (author	What you modestly imply is mere administrative data actually describes a rather staggering amount of work and for that you must be congratulated. Nevertheless, it would seem that the rather impressive amount of data gathering and crunching you've done has exhausted you – and you have omitted the next vital step, that of communicating and explaining the material.

response in bold)	1. Clear language follows from clear thought and at this point your article is not at all clear. There are far too many acronyms which you assume the reader understands; too many sentences that read poorly and details that are irrelevant or should be in a footnote or appendix (perhaps replacing Figure 1 "recruitment flow chart' that seems superfluous as you've explained the notion more than adequately in the article). And there are far too many acronyms: This reviewer did remember that "QI" means quality improvement but it took her a day to remember the "C" stood for continuous. Many readers will not even know that. From the first sentence of your abstract with "IHI" (only explained in reference 21) to FHT, QIIP and LC and more, this article is difficult to read and some sentences verge on the incoherent. On page 7 para 2 where you describe Tricco et al's review you describe the "learning collaborative model" and "taxonomy of interventions" which make little sense to anyone who has not read the original article. Lower down on the same page and on page 8 your description of your research as "conducted to explore the population level of the QIIP LC program ...." which is perplexing as by this point most of us have forgotten what QIIP or LC refer to. In short, you need to spend at least a fraction of the time you spent on the datasets on the writing and analysis. 2. Clarification is also called for with the health teams you refer to. What do these consist of? Family physician plus .. who? Physiotherapists? Nurse practitioners? Pharmacologists? Different jurisdictions define health teams differently and the reader is not necessarily familiar with Ontario's various primary care initiatives. 3. At times you focus on details that don't matter, for instance, SAS13.1 may well be a spectacular program but there is little point in including it outside of a footnote. 4. More important, the assumptions you make about the data seem problematic as your data does not indicate large differences between the controls and intervention groups. Your assumption that these health teams and other policy changes constitute "improvements" therefore seem premature if not inaccurate. As nearly as I can make out all that these various changes have resulted in is slightly better diabetes followup care; more prescriptions for statins (not necessarily a good thing) and ACE inhibitors; more colonoscopies and, basically, no change in visits to the ER or other clinical settings. (Barring an increase in ER visits for patients with COPD, ironically an acronym which you do explain, which to me implies that there may have been an element of iatrogenesis leading to patient stress and/or hypervigilance.) Certainly it seems that interpreting the data as positive, which you do ("outcomes were improved compared to controls", page 12), is not appropriate, for want of a better word. Data is not knowledge. It is the researchers who provide the context and possible explanations for what the data means. Here you need to reflect a bit more on what the data you have collected might imply for primary care. As someone who cares deeply about primary care I hope you succeed. **I was not able to make any additional changes based on the feedback from Dr. Baxter other than those made in response to the editors or other reviewers. I feel that the reduction in use of abbreviations and other clarifications already made address many of these.**
Reviewer 3	Mr. Bruno D. Riverin PhD
Institution	Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Que.
General comments (author response in bold)	Overall appreciation: The authors of this manuscript present the results of a controlled before-and-after study on the association between the implementation of a quality improvement program (through 'learning collaboratives' program) within Family Health Teams in Ontario and process of care and health utilization measures. The research question is relevant, and the choice of study design and analysis is appropriate; although I believe there could be improvements in how results are analyzed. The Methods section, however, is extremely hard to follow and needs major re-working to improve clarity. The results as presented appear valid, with a few modifications suggested below that would greatly improve the quality of reporting. In my opinion, the authors understate the risk of bias in this study; or fail to represent how limitations might have impacted their estimates. The manuscript as a whole is well-written. General: 1. Too many abbreviations are used which make reading more difficult; I recommend keeping these to a minimum. 2. Use controlled before-and-after study consistently (not pre- post-); including in the tables. Abstract: 3. P4; L11: define CQI, IHI; as a general comment. 4. P5; L12: do not use 'controlled evaluation'; mention the study design (i.e., CBA) Introduction: 5. I cannot comment on the comprehensiveness of the literature review. It is unclear, however, what the evidence of this research will add to the existing body of literature. I would have liked to have more information about the intervention; I am still not clear what the intervention is exactly; people outside of Ontario will likely not be too familiar with this either. Please describe the intervention in more details. 6. P7; L37-L47: It is very strange to read that 'controlled' studies are needed right after reporting on 4 clustered RCTs. More effort should be made to place this study within the body of evidence; i.e., how does this study improves on previous research? 7. P7; L42-47: 'as outcome improvements in prior studies…' I am not sure what this means and how is this linked to 'controlled' studies. 8. P7; L52: Remove the sentence ending in 'described elsewhere' and summarize the most relevant information in 1 or 2 sentences. What were the primary outcomes? Even if the results were not significant, it does not mean that they may not be relevant. 9. P8; L3-L8: The supplementary analysis of administrative data should be better explained within the larger study to make sure that the current research is not just looking for 'significant' results. Methods: This section requires major changes to improve clarity. 10. I find that the ordering of the text and of the subsections makes it difficult to follow. I suggest that the text follows the Flow Diagram (i.e., physicians, exclusions, patients included). Then I would follow with describing the study design, study period (i.e., when the program was implemented, what dates are used for 'before', what dates are used for 'after'). After this describing the data source seems appropriate. 11. P9; L22: Study design should be mentioned earlier in this section'; ie, move 'A controlled….' 12. P9; L25: Sentence starting with 'All datasets are linked…' seem to repeat information from L3. 13. P9; L30: What are the 3 waves? Index dates? This is not clear and should be made clear earlier. 14. P9; L32: I find the use of 'cross-sectional' and 'longitudinal' to describe the data is confusing and misleading; please use 'baseline', or data for the before-and-after periods. It should be made clear then 15. P9; L32-L48: Baseline characteristics and measure definitions should be presented with a minimum of information to improve reproducibility.

	16. P9; L57: Adjustment for baseline value? Why not use the difference-in-differences estimator? Why not use propensity score methods with characteristics before the intervention? I am not convinced that the best possible approach is used here. 17. P10; L6-L8: Was there any effort to account for clustering at physician/practice level? Results: 18. I wonder why confidence intervals are presented around the before and after estimates, and a p-value is presented for the difference; this really makes no sense to me. Why would you want to perform a statistical test on the values before and after separately? I suggest removing CIs around the before and after estimates, and estimating CIs around the difference; I do not find p-values useful here. CIs around the 'Adjusted Change' would, however, give some sense of the variability of the estimates; and also provide information about statistical significance. Referring to my comment above, I wonder if CIs are adjusted for clustering? Interpretation: 19. Changes needed for improvement. Overall, I am still not sure how the results from this study compare to those of other studies; particularly in terms of quality of evidence. 20. P12; L27: 'These differences were….' is not clear, and possibly not appropriate; Do you mean that your sample was sufficiently large to detect differences? 21. P12; L32: 'It should be noted….'; Because you have population-based data, you could provide population numbers instead of alluding to a population impact. 22. P13; L30: 'This hightlights one of the advantages…'; This statement is incorrect; population-based data is not more complete; it simply covers the whole population. Remove statement. Also, I am not convinced that the data 'coverage' can account for the difference in results across studies. Please clarify. Limitations: 23. This section does not cover the major limitations. 24. P14; L40: I am not sure that the limitations of using administrative data are explained sufficiently; and how this could impact the results. 25. P14; L52: 'This was a program implementation…"; how does this impact the results; bias? 26. P15; L3-L8: 'Finally…'; would those other reforms be implemented differentially across groups? If so, only controlling for time-trend would not necessarily be sufficient. 27. P15; L41-L43: 'This is likely because…'; this should probably be addressed in the limitations, not in the conclusion. Also, replace 'outcomes' by 'measures'. **Most substantive comments already addressed in the revised methods and tables in response to comments from the editor or Dr. Levitt.** **We have adopted consistant before and after study design language throughout the paper and in the tables. Comments 8-9 have been addressed by a change to the final section of the introduction. The methods section has been revised to incorporate his suggestions (comments 10-15).** **Comment 16 - In terms of choice of analytic methods we agree that the other methods he suggests (difference in difference or propensity score matching) were and are viable alternatives which we considered. We feel the methods presented are equally valid and are similar to those that we have used in many other studies using these data sets. As the audience for this paper is practicing physicians, other primary care practioners and health systems managers who may not be as familiar with those methods we felt this approach would be more easily followed by our primary target audience for this paper.**