Reviewer comments

Article ID: CMAJ Open-2022-0006
Article title: Effect of a multimorbidity intervention on health care utilization and costs in Ontario: RCT and propensity-matched analyses
Authors: Bridget L. Ryan PhD, Luke Mondor MSc, P. Wodchis PhD, Richard H. Glazier MD MPH, Leslie Meredith MEd, Martin Fortin MD MSc, Moira Stewart PhD

Reviewer 1: William Pace / University of Colorado, Family Medicine

The comment in the first paragraph of the Results section page 4 lines 53 -54 concerning "no statistically significant differences in baseline characteristics" should be removed. As this was a person level RCT by definition any differences would be random. Thus, p values should not be commented upon for baseline data. They are correctly not included in Table 1.

We agree and, in the Results, RCT Analysis have deleted the sentence, which stated, "There were no statistically significant differences in baseline characteristics between the groups."

We have adjusted the RCT analysis section. Please note we had pre-intervention in wording for the RCT methods and indicated a generalized estimating equations had been used, and this was incorrect. The analysis for the RCT now reads, "For the RCT analysis, differences in health care utilization one year after index dates for PACE intervention group versus control group (control used as reference group) were estimated using univariate regression with generalized linear model (GLMs) including a single covariate for RCT assignment (intervention or control)."

The analytical methods and Table 2 output need to be better matched.

While the entire study is plagued by a very small sample size, given that the analysis between intervention and RCT controls demonstrated worse outcomes in the intervention arm it is hard to see how a larger sample size would demonstrate significant differences. This part of the limitation section should be reconsidered.

We agree and have removed the first paragraph of the Limitations section where we referenced large sample sizes as a limitation overall.

We have added to the first paragraph of the Interpretation section that, for the outcome 7-day follow-up with family physician after hospitalization, which showed a trend in favour of the intervention, "A larger sample size may have permitted us to see a statistically significant difference. The difference found in this study can suggest clinical and policy importance and provide hypothesis generation for future studies."

Overall, it is hard to see how the reported outcomes add significantly to the current literature in the area.

As noted under the Editor's Item #1, we identified only six evaluations of interventions for patients with multimorbidity in the literature. Only two of these evaluated health care utilization as we did in this paper. These two evaluations found similar results to ours; that is, no significant effects of interventions for person with multimorbidity. It is important to publish this Canadian work to add to the limited international evidence.

Of note the original protocol did not specify these exact same outcomes for analysis…nor did it discuss how multiple comparisons would be handled. This is perhaps less of an issue since all of the findings are negative.
But, in theory, converts the different outcomes to exploratory from secondary and would disallow reporting of p values due to the lack of specificity in the protocol and the exploratory nature of the outcomes. As this reviewer did not have access to the statistical analysis plan perhaps this was clarified prior to data lock.

As noted in the Editor's Items [not included] above, the health administrative outcomes, unfortunately, were not listed separately in the "*Original Secondary Outcome Measures*" section on ClinicalTrials.gov (#104191); however, they were always intended as Secondary Outcomes as stated as part of the "*Descriptive Information*" under section "*Detailed Description*" in the clinical trial registration and as stated in the Protocol paper.

Additionally, we note in the manuscript, "Minor adjustments from protocol were made to some outcome definitions to align with measures available in health administrative data". As an example, avoidable hospitalizations was only available for people under 75 years of age. This would have excluded many of our participants and so the change was made to acute hospitalizations.

We did not make adjustments for multiple comparison, but we conducted sensitivity tests to assess robustness. Also, we conducted both an RCT analysis and a propensity-matched analysis with consistent results. As the Reviewer notes, this is less of an issue, given there were no significant findings.

Consistent with our responses to the Editor's Items #6 and #8 above, because these outcomes were published a priori in the protocol paper and in the "*Descriptive Information*" under section "*Detailed Description*" in the clinical trial registration, we believe it is reasonable to report p-values.

Reviewer 2: Imaan Bayoumi / Queen's University, Family Medicine, Kingston, Ont.

Additional details on the trial would be helpful. The sample size is quite modest for an evaluation of this complexity, …
Please refer to our response to the Editor's Item #2 [not included].

…and I presume there were recruitment challenges, but it would be helpful to better understand why the sample size was small and whether the target sample size was achieved.

Please refer to our response to the Editor's Item #7. [not included].

There were no significant differences in any of the 5 selected outcomes for the RCT population or the propensity score matched cohort.
The authors are to be commended for submitting a negative trial.

Thank you. Given the paucity of evidence on complex interventions for people with multimorbidity, we agree it is important to publish this work.

They highlight the small sample size as an explanation for the null finding but also note that it is consistent with the results for other similar interventions.

We agree this is confusing and as indicated in our response to Reviewer 1 Item #4, we have removed from the Limitations section the sentence, "The main limitation for this analysis was the small sample size of 82 intervention participants and 74 controls. This may explain the lack of statistical significance found in the RCT analysis."

This removal necessitated a slight rephrasing in the first sentence of the third paragraph of the Limitations, where it now states, "In the propensity-matched analysis, for propensity-matched controls, it was not possible to match for every baseline characteristic collected in the questionnaires because these individual-level characteristics are not available in the health administrative data and so could not be included."

Perhaps the intervention itself is not robust enough to affect health services outcomes
The outcomes selected are also all selected with a policy maker lens, rather than a patient lens. Perhaps primary care visits or continuity of care may have been additional meaningful outcomes to examine using health administrative data.

We discuss this in the Limitations section and have now made this more explicit, and have added that primary care utilization may be meaningful to consider (changes underlined here), "Our outcomes were limited to hospital care and direct costs of health care services, outcomes relevant to policy makers. The intervention may confer benefits aligned with outcomes that are situated in primary care such as continuity of care or that consider patient preferences such as improved function but not confer benefits aligned with health care utilization and costs."