

Article ID: 2021-0265

Article title: Identifying subgroups of adult high-cost health care users: a retrospective analysis

Authors: James Wick MSc, David J.T. Campbell MD PhD, Finlay A. McAlister MD PhD, Braden J. Manns MD PhD, Marcello Tonelli MD SM, Reed F. Beall PhD, Brenda R. Hemmelgarn, MD, PhD, Andrew Stewart MSc, Paul E. Ronksley PhD

Reviewer Comments and author responses

Reviewer: 1

1. The article seems to contribute to what is already known on this topic. However, the authors need to do a better job in explaining their contribution to the literature. They refer to the Anderson et al (2019) study and indicate that it used a two-year period to define high-cost patients. P. 4, l. 25-28: "A recent Canadian study of high-cost patients described the clinical and demographic profiles of the top 10% of individuals by cumulative cost over a two-year period." Yet, they criticize this study (along with another one from the US) for defining their population at a single point in time. P. 4, l. 41-44: "These two previous studies defined their populations at a single point in time (i.e., episodic healthcare use) rather than those that are high-cost in multiple subsequent years..." This seems confusing and the authors need to clarify this point. That's a very important precision to make to accurately represent current knowledge in this field and to determine the extent of the contribution of this manuscript.

Thank you for highlighting the need to clarify the novelty of this work. We've reworded a section of the background to make it clear that the prior studies defined their cohorts using a single period in time (a two year period in Andersen's case) rather than persistence over separate independent time periods.

Introduction Line 13-15:

"A recent Canadian study of high-cost patients described the clinical and demographic profiles of the top 10% of individuals by cumulative cost accrued during a two-year window"

2. It would also help if the authors could briefly explain the CIHI population grouping methodology that was used in the study by Anderson et al. To what extent it is similar or different from the latent class analysis.

The CIHI Population Grouping Methodology uses diagnoses from health care encounters in various settings to summarize diagnosis codes into a clinically meaningful set of 226 health conditions. Detailed description is available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.191297/-/DC1. The CIHI grouping methodology was not designed to predict high-cost cases, and previous work has already shown that the model performs better for low- and moderate-cost users than for highest-cost users. Further, the population grouper is a method for summarising data into pre-determined groups, whereas latent class analysis is a statistical technique classifying information into a data-driven number of groups based on its characteristics. Thus the two methodologies are distinct.

For purposes of brevity in the manuscript, we have not provided an exhaustive comparison of these methodologies.

3. The authors haven't added an explicit research question but have stated a clear objective in the abstract: " We sought to identify homogeneous groups of persistently high-cost patients to inform future interventions." They need to repeat this in the body of the manuscript and elaborate a bit more.

As suggested, we have made the last sentence of the introduction an explicit objective and descriptive of the process taken.

4. P. 7, second paragraph: Reference to specific table(s) would help. Table 1, eTable 1?

Thank you for the suggestion. We have now included further information on the initial 21 classes and how they correspond to the 9 final classes, and thus have referenced this new table (eTable 2) within the results section on page 7.

5. P. 8. L. 48: Inflammatory bowel disease—IBD is used later. So, the acronym should be explicitly mentioned here.

We have defined the IBD acronym in the noted place.

6. P. 9, L. 9 -13: Not obvious to link these numbers with those in Fig. 3. Make it more explicit. And giving a title to this figure and the other ones will help.

Thank you for noting this error. The in-text reference to Figure 3 is now moved to its appropriate place at the end of the paragraph. The figure legend, including title of Figure 3, was included on a separate page per CMAJ formatting guidelines.

Reviewer: 2

Introduction

1. *“Further complicating this goal, two-thirds of the high-cost population in a given year do not remain high-cost in the following year, making it difficult to identify the appropriate target population for intervention.”*

· This is true but it may also be the case that some individuals may return to the high-cost state in the following year; examining patterns of high-cost over time may be helpful.

We agree that there are a number of iterations of cost patterns not accounted for in this sentence as originally stated, thus we have modified the statement to account for broader scenarios. We now state: “Further complicating this goal, two-thirds of the high-cost population in a given year are not high-cost in the following year, making it difficult to identify the appropriate target population for intervention”

2. *“A recent study of a US Kaiser Permanente Northern California-based cohort used latent class analysis to identify 7 subgroups of medically complex patients.*

However, their classification scheme was developed in a subset of patients with high comorbidity and likelihood of acute care use as opposed to the general population.”

• **Unless something is missing here, this is ultimately the approach the authors take in this analysis. Yes, they start with the general population, but they then narrow it down to a subset of patients who also have high comorbidity and high likelihood of acute care. It is not clear how exactly this differs from what the authors have done here. Further clarification would help.**

Thank you for the opportunity to clarify. While there is likely to be some overlap between a source population of 'high comorbidity and likelihood of acute care use' and a subset of the general population based on cost, the two groups are not synonymous. Within our population, there were various levels of acute care use and comorbid status (as shown in Table 1, and Figure 3). This would not have occurred had we restricted our cohort to medically complex patients with high likelihood of acute care use. Thus, while the two methods each result in a cohort of costly, generally complex individuals, there are substantial differences.

Methods

3. “*The presence of comorbidities was ascertained using validated algorithms within the administrative health data using ICD-9/ICD-10 coding*”

• **Which chronic conditions are captured by these algorithms? These are only stated later on.**

We have modified this sentence to detail that these algorithms define 13 chronic conditions. However, we feel that a list of each of them would be disruptive to the flow of the paper, given they are displayed within Table 1.

• **It would be useful to state the data sources used to obtain utilisation and cost data (e.g., DAD, NACRS, etc.).**

We have included information about the data sources to obtain utilisation and cost data within the section titled ‘Data Sources and Selection of Classifying Characteristics’.

• **Were day surgeries captured in the DAD or NACRS? If the former, were these included in hospitalisations?**

Day surgeries were captured within the Ambulatory Care Classification System file (Alberta’s version of NACRS)

• **Who is captured in the Pharmaceutical Information Network data? Does this capture everyone, regardless of the type of drug plan (public versus private)? If the latter, is it in line with the perspective of the cost analysis?**

The Pharmaceutical Information Network data captures medication prescription for everyone in the province, regardless of drug plan. Private insurance doesn’t preclude an individual’s inclusion within the PIN data.

• Do outpatient physician encounters only include physician visits?

Outpatient physician encounters included all encounters between a physician (GP or specialist) and a patient that occurred in an outpatient setting. Inpatient and emergency department encounters were not captured within this specific data source, rather, included in the Hospitalization and ED sources, respectively.

4. “For hospital and emergency department/hospital-based ambulatory care encounters, the resource intensity weight of the encounter (an estimate of the relative cost associated with elements of a given encounter, including demographic characteristics, diagnoses, and procedures) was multiplied by the mean provincial cost of a typical encounter.”

• Was this also the approach taken to estimate costs of psychiatric hospitalisations, which tend to have long lengths of stay?

We took this approach for all hospitalizations, including psychiatric hospitalizations which are associated with a longer length of stay. While we didn’t manually account for psychiatric hospitalisations, the Resource Intensity Weight for such hospitalisations accounts for the expectation of extended length of stay.

5. “All costs were estimated using information for the fiscal year in which they took place.” · **Are costs in nominal dollars then? Why were costs not inflated to the latest year (that is, expressed in real dollars)?**

We received a similar comment from Reviewer 3 (Comment 2). We thank you both for the comments and feel it is appropriate to use nominal dollars for identifying high-cost users in any given fiscal year and persistent high-cost users across multiple fiscal years. However, where we reported averages or cumulative costs across several years, we have now adjusted those costs to reflect 2018 dollars.

6. “Each individual’s cumulative cost for all healthcare encounters within each fiscal year was calculated, and those in the top 1% of cumulative spending were defined as ‘high cost’. An individual was defined as ‘persistently high-cost’ if their annual cumulative spending was in the top 1% of costs for at least two consecutive fiscal years between 2014 and 2019.”

• Why was the top 1% chosen and not the top 5% or top 10%, given that other work have used these cutoffs as well? What are the pros of your approach? So, to clarify, were persistent high-cost patients defined as those in the top 1% of the cost distribution for 2, 3, 4 or 5 years and non-persistent high-cost patients (not examined here) defined as those in the high-cost state for only 1 year? How was this definition determined?

As also discussed in Editor Comment 1, while there are several common ‘high-cost’ cutoffs in the literature, including top 5% and top 10% of costs, this includes literature that has used 1% cut-points (e.g. Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. CMAJ 2016;188:182-8). For the purposes of this study, we decided on a 1% cut-point to

isolate only the most consistently expensive patients, allowing for a focused population for latent class analysis, whereas 5% and 10% cut-offs may include more 'noise'.

The reviewer is correct in that ‘persistent high cost’ was defined as the top 1% of cost distribution for at least two (and up to five) consecutive years, and ‘episodic high cost’ (or non-persistent) those for only one year. This definition has been used in other papers in the field (e.g. Figueroa JF, Zhou X, Jha AK. Characteristics and spending patterns of persistently high-cost Medicare patients. *Health Affairs* 38(1):107-114; Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. *CMAJ*. 2016;188(3):182-8)

• Did the authors examine patterns of high cost over time? This may help inform how persistent high-cost patients are defined.

While an interesting area of work, we did not examine patterns of high cost status over time as we felt this was outside the scope of the current manuscript, and has been done previously, including Ronksley et al. (Ronksley PE, McKay JA, Kobewka DM, Mulpuru S, Forster AJ. Patterns of health care use in a high-cost inpatient population in Ottawa, Ontario: a retrospective observational study. *CMAJ Open*. 2015 Jan;3(1):E111.)

• Can one assume that individuals in the high-cost state for 2 years are similar to those in the high-cost state for 5 years? Previous research on this topic has shown that these individuals can be quite different – see “Figueroa JF, Zhou X, Jha AK. Characteristics and spending patterns of persistently high-cost Medicare patients. *Health Aff (Millwood)* 2019;38:107-114” and “de Oliveira C, Mason J, Kurdyak P. Characteristics of patients with mental illness and persistent high-cost status: a population-based analysis. *CMAJ*. 2020 Dec 14;192(50):E1793-E1801.”

As also discussed in Editor comment 4, we feel that including a definition that spans 2, 3, 4 or 5 years to define ‘persistence’ allows for capture of different types of high-cost patients rather than restricting our cohort and potentially negatively influencing our findings. We aim to capture groups who’s high spending is based on similar health issues, though not limited to similar temporality of high-cost.

We have provided the distribution of consecutive years of high-cost across the nine final groups below. While there is some minor variability across groups, the results suggest the groups are not materially different, as evidenced by the mean years of high cost being within 6 months across groups (bottom row).

# of years	Severe Mental	Unstable Housing/ Subs abuse	Advanced CKD	Cancer	Cardiovascular- related	Biologics	Rehab/care after surgery	COPD/Resp	Dementia/comm place	Total
------------	---------------	---------------------------------	--------------	--------	----------------------------	-----------	-----------------------------	-----------	------------------------	-------

2	51.4	47.0	49.4	67.7	60.5	48.0	71.3	64.3	75.0	59.8
3	31.2	29.9	24.1	19.8	23.1	25.5	20.2	23.4	20.1	24.0
4	11.8	15.5	11.8	7.6	8.4	14.2	6.2	7.6	3.4	9.2
5	5.5	7.7	14.7	4.9	7.9	12.2	2.3	4.7	1.6	7.0
Mean years	2.71	2.84	2.92	2.50	2.64	2.91	2.40	2.53	2.32	2.63

If the editors desire, we would be happy to include this information in Table 1 to provide more information on the relative contribution of years of persistent high cost to the 9 groups.

7. “Using the classifying characteristics, we began by specifying a model with two classes (representing subgroups) and added classes iteratively until the Akaike Information Criterion and Bayesian Information Criterion were minimised, indicating the appropriate number of classes to maximise model fit while minimising the potential for overfitting.”

• Did the authors have a pre-hypothesised number of potential groups based on previous literature and opinion expert, or was this a merely data driven exercise determined by the values of the AIC and BIC? This is later explained but it may be useful to provide more details earlier on.

The number of groups was exclusively driven by data. While we were aware of a number of different groups defined in prior literature, we weren't bound by these numbers. We took the data-driven groups to experts to combine based on similarities in clinical management and compared them to prior literature to ensure face validity.

Results

8. Figure 1 is a bit confusing and not particularly useful; it's not clear how the persistent high-cost patients were chosen based on the number of high-cost individuals.

While we appreciate the comment, we feel that figure 1 has value in demonstrating that the study population is based on a 1% cut of the Alberta population in 5 separate years, and that the persistently high cost population is a subset of these 1% cuts.

9. The figure in Appendix C provides a bit more information, but still requires some clarification. For example, is it correct to say that 3,909 individuals were high cost patients in 2014/15 and 2015/16 and that, among these, 1,280 were high-cost patients in 2014/15, 2015/16 and 2016/17? Is it the case that 152 individuals were high-cost patients in all 5 years?

The segments of the Venn Diagram are mutually exclusive, such that each individual belongs to only one segment, and the entire figure sums to 100%. To the example referenced here, the correct interpretation is that 3,909 were high-cost in 2014/15 and 2015/16 (though not subsequently). In addition 1,280 were high-cost in 2014/15, 2015/16, and 2016/17 (though again, not subsequent years). There were 1,482 persons who were high cost in all years (we intend the darker shading to assist with distinguishing overlap of segments), while 152 patients

were high-cost in 2014-15 & 2015-16, and 2017-18 & 2018-19 (but not in the intervening years).

We feel that this ‘exclusivity’ of regions is typical of such figures, and thus propose not changing the figure, however we are happy to defer to the editors on whether clarification/modification of the figure is appropriate. .

10. Do the authors have data on the 3,909 and 152 individuals? It is likely these are very different groups of individuals, yet both are considered high-cost (see comment above).

As mentioned in the response to a similar query above, we didn’t specifically explore the differences between individuals with different lengths of persistent high-cost status. However, we did explore the distribution of years of high cost across the nine high-cost group.

We have provided the distribution of consecutive years of high-cost across the final nine groups below. While there is some variability across groups, the results suggest they are not materially different, as evidenced by the mean years of high cost being within 6 months across groups (see bottom row of the table).

# of years	Severe Mental	Unstable Housing/ Subs abuse	Advanced CKD	Cancer	Cardiovascular- related	Biologics	Rehab/care after surgery	COPD/Resp	Dementia/comm place	Total
2	51.4	47.0	49.4	67.7	60.5	48.0	71.3	64.3	75.0	59.8
3	31.2	29.9	24.1	19.8	23.1	25.5	20.2	23.4	20.1	24.0
4	11.8	15.5	11.8	7.6	8.4	14.2	6.2	7.6	3.4	9.2
5	5.5	7.7	14.7	4.9	7.9	12.2	2.3	4.7	1.6	7.0
Mean years	2.71	2.84	2.92	2.50	2.64	2.91	2.40	2.53	2.32	2.63

If the editors desire, we would be happy to include this information in Table 1 to provide more information on the relative contribution of years of persistent high cost to the 9 groups.

11. “Using latent class analysis, we initially identified 21 potential subgroups of persistently high-cost individuals based on 20 classifying variables. Consultation with our clinical experts resulted in a reduction from 21 to 9 clinically meaningful persistent high-cost subgroups based on similarities in patient characteristics and clinical needs.”

- Yes, exactly, this is the point made earlier. One can do this exercise for a long time until one reaches the lowest AIC and BIC values but can then end up with a large number of groups, which can be challenging to interpret. It would be worthwhile to make this clearer earlier on.

We understand the point the reviewer makes, however based on the sequence of events described within the methods and results, we have kept the sections as they are as we feel this provide a clear delineation of the steps used within this analysis.

12. Figure 2 provides a nice summary of the final groups chosen but does not mention the 21 groups obtained and how these were boiled down to 9 groups. It would be good to have a bit more information on how these 9 groups were defined. Also, did the authors provide data on the latent class analysis? I don't believe this was included in the manuscript.

As also suggested by Reviewer 1 (Comment 4), we have added a supplementary table (eTable 2) to provide information on the initial 21 classes, and how they were collapsed into the nine final classes.

13. For the group 'Treatment for cancers (3.9%).' was this group solely determined by whether an individual received chemotherapy, radiation therapy or cancer directed surgery in any of the years in which this individual was in the high-cost category?

- **Did these authors have access to cancer registry data? This could have been helpful to determine whether someone had cancer or not.**
- **4% seems a bit low given that cancer care is typically very high.**

The identity of this group doesn't indicate that only 4% of high cost individuals have cancer, rather that of the persistent high cost population, 4% were predominantly defined (based on costs and characteristics) based on cancer. As these classes are not mutually exclusive, patients with cancer were present in the other 8 groups, but their profile was best defined by something other than their cancer status. While cancer registry data would have been helpful, this was not available to our research team. However, cancer as a comorbidity was defined based on validated algorithms with strong sensitivity and specificity, and cancer medications were identified from PIN data suggesting that misclassification of this patient population was unlikely.

14. *The persistent high-cost population (n=21,115) was responsible for \$2.7 billion (CAD) in healthcare spending annually (approximately 13% of the annual Alberta healthcare budget in the timeframe of this study)...*

- **Is this value for the 5 years or on average? What year are the 2.7 billion dollars in? Is this value for all 5 years as well? In nominal or real dollars?**

Thank you for the comment. This value reflects average annual cost and has been updated to reflect annual cost in 2018 dollars. The correct value is \$2.8 billion and the revised manuscript has been updated to reflect these changes.

15. *"...and subgroup costs ranged from \$105,000/person/year in the subgroup of patients receiving biologic therapies for autoimmune conditions to \$174,000/person/year among patients experiencing severe mental health conditions."*

- **Information on Figure 3 requires more clarification. What year are these costs for? Are these total costs across all 5 years? Although one can calculate the mean cost from the values in the figure, it may be worthwhile including the mean values too.**
- **Also, the authors may want to include other measures, such as the median and 95% confidence intervals.**

The per person/year costs have been updated to reflect 2018 dollars. We have also included mean and median costs within eTable 3 (as per suggestions from the editors and reviewer 3).

16. “Hospitalisations accounted for 48.3% of overall cost, while 26.5% of overall cost was attributable to medication use, 15.0% to ED encounters, and 10.2 % to outpatient physician encounters.”

- **For the group with severe mental illness, do hospital costs include both acute and psychiatric hospital costs? It may be worthwhile to separate these two, in particular for this group.**

While we recognise the psychiatric hospital costs may be higher based on longer lengths of stay as referenced in Reviewer 2's previous comment, we felt that exploring the types of hospital spending relevant in only one (or a small number) of groups, distracts from the larger point of identifying groups within a heterogeneous population. Exploring types of costs relevant to specific subgroups represents an interesting avenue for further research within each of the nine subgroups.

17. “Age varied across the 9 subgroups (ranging from a mean age of 41.8 years to 81.1 years, with an overall mean of 60.2 years, SD: 18.4 years).”

- **It would be helpful to state which group for each of the values.**

Thank you for the suggestion. We have updated the text so that the mean ages mentioned above are referring to a specific high-cost group.

18. “The most prevalent, and costly diagnoses of hospital, ED, and physician encounters, as well as prescribed medications were tabulated for each subgroup (Online Supplement A).”

- **Do the authors mean eTable 1? It might be helpful to mention some of these in the manuscript.**
- **What do the authors mean by cumulative cost? Do they mean the cost over the 5 years?**

Thank you for pointing out the error, the reference to Online Supplement A has been changed to eTable 1. We appreciate the suggestion to mention these within the manuscript, however we decided not to include additional text given word count limitations. However, if the editors feel this an important inclusion we are happy to make this revision.

Cumulative cost refers to accumulated cost over the period individuals (and by extension groups) were persistently high cost.

Interpretation

19. *“Previous literature has identified specific characteristics and healthcare utilisation associated with high-cost patients, but have not considered persistent high-cost users, an important limitation given that the majority of high-cost users in any one year are not high-cost in subsequent years and so intervening is less feasible.”*

· I don't believe this statement is correct; see “Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. *CMAJ* 2016;188:182-8,” “Figueroa JF, Zhou X, Jha AK. Characteristics and spending patterns of persistently high-cost Medicare patients. *Health Aff (Millwood)* 2019;38:107-14” and “de Oliveira C, Mason J, Kurdyak P. Characteristics of patients with mental illness and persistent high-cost status: a population-based analysis. *CMAJ*. 2020 Dec 14;192(50):E1793-E1801.”

We agree the original statement is not correct and have clarified the text to more accurately represent our intended meaning.

20. **How were the interventions described in Box 1 determined? Some information on this would be helpful.**

We have added a line to the ‘Expert Opinion’ section of the methods to specify that this was completed by the expert panel and guided by literature review.

21. *“Second, inpatient encounters tend to be costlier than community-based alternatives, and thus interventions which decrease reliance on hospital services may be promising in all subgroups.”*

· **How feasible is this and for which groups is this feasible? It sounds like this is a possibility for high-cost patients with CVD but perhaps not for all high-cost patients. Although not possible to examine all cases, it may be useful to examine those with the highest costs or for which small changes can provide the biggest impact.**

· **For example, for severe mental health conditions, the target of the intervention is the high cost of inpatient treatment (do the authors mean psychiatric or acute inpatient treatment or both?) and the type of intervention is around reducing inpatient services. The authors suggest preventative treatment of psychoses, and minimising the need for institutionalisation and inpatient care. However, this is not straightforward. There is no proven way to prevent psychosis. Moreover, according to de Oliveira et al. (2020), “some psychiatric hospital admissions are likely part of the natural course of the disease, and thus inevitable”; in addition, the authors state “some admissions to hospital, namely readmissions, could potentially be prevented through timely access to physician care and adherence to medication.” (Incidentally, at first blush, the 2 references provided for severe mental health disorders do not seem to be very useful or directly related to severe mental illness.)**

Thank you for these salient points. Box 1 is intended as a starting point for consideration of potential interventions in these groups, without speaking to important questions such as the feasibility of the interventions, what elements of local context would drive their success, barriers

and facilitators to success, and so on. We have added a sentence to this portion of the interpretation to highlight that these previously identified interventions are worthy of consideration, rather than an implied 'ready-to-implement' fix.

We agree with the reviewer that the quoted statement would require substantial explanation and evidence-based defence, beyond the intent of its inclusion in the paper. We appreciate the points made by de Oliveira et al., which are also potentially relevant to other subgroups, and have thus modified the quoted statement to emphasise that decreased reliance on hospital services *where viable* is a worthy aim, acknowledging that not every inpatient encounter is avoidable.

With specific regard to the proposed severe mental health intervention, we have modified the wording to indicate the ideal for upstream detection and intervention, rather than prevention, which we agree is not the correct term for this population. Thank you for drawing our attention to the recently published de Oliveira paper, which we feel represents strong, recent evidence into the care for this population. We have modified the suggested intervention and provided new references accordingly.

22. Have the authors also examined the proportion of preventable acute care among these patients, as examined in the previous work by this group? See Ronksley, P.E., Kobewka, D.M., McKay, J.A. *et al.* Clinical characteristics and preventable acute care spending among a high cost inpatient population. *BMC Health Serv Res* 16, 165 (2016). <https://doi.org/10.1186/s12913-016-1418-2>. This information could also provide options to help curb costs.

This is an interesting avenue to explore further. This has been previously explored within one of the classes (Advanced CKD - Chong, C., Wick, J., Klarenbach, S., Manns, B., Hemmelgarn, B. and Ronksley, P., 2021. Cost of Potentially Preventable Hospitalizations Among Adults with Chronic Kidney Disease: A Population-Based Cohort Study. *Canadian journal of kidney health and disease*, 8, p.20543581211018528.), and exploring preventable spending in these classes represent an area for future work. Thank you for the suggestion related to the BMC Health Services paper. This paper was used to help identify relevant interventions for the 9 classes.

Limitations

23. “History of homelessness was defined based on the presence of the ICD-10 code Z59.0 in hospital and/or ED records in the prior year, or 45 days subsequent to the index date.” This algorithm will only capture individuals who access hospital-based care; it may be worthwhile stating the limitations of using this algorithm.

We agree this is a relevant limitation of the algorithm and we have now referred to it in the limitations section accordingly.

Minor comments

24. “Further, these two previous studies defined their populations at a single point in time (i.e., episodic healthcare use) rather than those that are high-cost in multiple subsequent years, who arguably have greatest potential for targeted intervention and cost savings.” Is there a reference for this statement?

We have now provided a reference for this statement. Billings J, Mijanovich T. Improving the management of care for high-cost Medicaid patients. *Health Aff (Millwood)* 2007;26:1643-54.

25. “severe mental health conditions (e.g. schizophrenia, psychosis) (14.5%)” Note that schizophrenia is a form of psychosis so there is some redundancy here; this could be changed to ‘schizophrenia, other psychoses).

We agree this was slightly redundant and have made the suggested change.

Reviewer 3

1. The strength of this study is in its use of health administrative data. However, it is important to note that the study did not include healthcare use and cost from all sectors. Specifically, as the authors have pointed out, they did not include long-term care costs. In Wodchis et al.’s analysis of high-cost users using administrative data in Ontario, nearly 40% of people in the top 1% did not have acute care utilisation. Among these patients, costs incurred in the continuing care sector (mainly attributable to long-term care) accounted for about half of all expenditures. This is a significant limitation in a study that aimed to develop a population-based typology that can inform clinically- and policy-relevant interventions for reducing healthcare costs. If long-term care costs were included, one might expect the groupings to look different from what the authors presented in the article.

We agree that LTC costs are important and may have provided further information in defining groups. We used proxies for long term care costs (including an indicator of long-term care residence) to account for this to the extent possible and believe the inclusion of additional data would not have meaningfully changed the groups we found. However, we have included a statement to this effect in the limitations section as suggested.

2. Secondly, for identifying high-cost users in any given fiscal year and persistent high-cost users across multiple fiscal years, using the unadjusted cost in each fiscal year is appropriate. However, when calculating and reporting averages or cumulative costs across several years (e.g., in the section on ‘Sources of Health Care Cost’ and eTable 1), it is necessary to adjust costs obtained from different periods for inflation (e.g., by indexing to the most recent year within the study period).

Thank you for this advice. As also suggested by Reviewer 2, we have adjusted costs to 2018 dollars where appropriate.

3. One minor comment regarding Figure 2. While Figure 2 makes a beautiful infographic, it is not a good graphical representation of data - in the sense that it did not offer any new data or novel interpretation of the data than what is already presented by the bulleted points on page 7, lines 38 to 56. I found the Venn diagram presented in Appendix C to be more informative and would suggest the removal of Figure 2 from the paper and move the Venn diagram as a main exhibit in the paper.

We agree that Figure 2 doesn't provide sufficient additional information to warrant inclusion. We appreciate the support for Appendix 2, however per the editors suggestion in comment 13, we have included a new Figure 2 as a way of visualizing differences between the classes and have left the Venn Diagram as an supplementary figure.