PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Social and Behavioral Factors Associated with Frailty Trajectories in a Population-Based Cohort of Older Adults
AUTHORS	Chamberlain, Alanna; St Sauver, Jennifer; Jacobson, Debra; Manemann, Sheila; Fan, Chun; Roger, Verinique; Yawn, Barbara; Finney Rutten, Lila

VERSION 1 - REVIEW

REVIEWER	Arnold Mitnitski Dalhousie University, Canada
REVIEW RETURNED	23-Feb-2016

GENERAL COMMENTS	This is generally thoroughly done study. The authors rightly mentioned that most studied examining social and behavioral factors were cross-sectional and just a few studies considered longitudinal changes. Some of existing studies escaped attention of the authors (Mitnitski et al. Exp Ger 2013) although without considering social and behavioral factors. One of the major finding is that these factors were stronger associated with frailty changes in younger group. No such association was found in the 80-89 year old group. This result might undermine (in my opinion) their conclusion that social and behavioral factors leading to increasing frailty may inform interventions for individuals at risk of worsening frailty, although the authors discussed this in the paragraph before the Limitations and Strengths section. I am not sure that the reference [32] in respect to '…can be constructed using deficits available in the electronic medical record' is correct (I have not found such a statement in [32]) although the frailty index indeed had been created using the electronic medical records by J Young and A Clegg and presented at the British Geriatric Society meeting(s)). I am not sure that I understand the last 3 sentences (lines 42-53). Figure is identical to just published in the J Am Geriatr Soci 2016;64: Ethics board review statement is missing.

REVIEWER	Phil St John University of Manitoba Canada
REVIEW RETURNED	24-Feb-2016

GENERAL COMMENTS	I can't do a full review since Figure 1 is missing, so I have checked the major revision box. As well, longitudinal data analysis is very complex and there are many ways to model these data, so a statistical review would be important. I cannot comment on the English, since I cannot find the journal format for US or UK spelling. There are relatively few prospective cohort studies with the data required to determine frailty trajectories, and fewer still with social and behavioural factors, so this study is a useful addition to the literature. Briefly , the authors constructed a Frailty Index (FI) from administrative data from a well known population based health care registry, and show that social factors are associated with frailty in prospective data analysis. In particular, education and concerns about alcohol misuse strongly predict worsening frailty. The paper is well written, easy to follow, and is fairly important. While not too surprising, the findings are important clinically and for policy. Major Points 1. I do not have Figure 1. 2. These analyses are based on administrative data. These are often not as accurate or carefully gathered as primary data. The authors should acknowledge this as a limitation. 3. There is considerable debate (and indeed disagreement) regarding the conceptual model and definition of frailty. The "Frailty as a Phenotype " model is quite different, and may have different trajectories. The FI was chosen (I suspect) because the data were available. Ideally, one would have both measures of frailty. The authors acknowledge this a little, but should comment a bit more on their choice of a measure of frailty. This is in no wise a fatal flaw, simply something which requires explanation. 4. There are quite a number of ways to analyse these data (none particularly better than the other), and it is very complicated. The paper should likely be sent for statistical review. An appendix on statistical techniques and whether the modeling assumptions were met would be a good idea. As well, the authors should likely do some sensitivity analyses (GEE and multilevel models, for instance) to see if the results are the same. If they are (which they likely will be), then it strengthens the argument. 5. The criteria for an item to be included in the FI (common enough to saturate a model, adverse, stable over time etc) also include age. So its difficult to comment on age-stratified analyses, since its really looking at the same predictor variable twice. It would be reasonable to stratify disability, multimobidity, and so on, but I am not sure what it means to stratify frailty on age. Of course, the question is relevant and interesting. Since its only one criterion, it may not have a large influence, and I would be inclined to simply ask for a statistical review of this issue. (It may be more reasonable to drop this as a criterion for inclusion in the FI). 6. In most countries, there has been a strong cohort effect of

education. So usually education and age are highly correlated. This is particularly true on the Great Plains, where an incomplete education in one room schools was the norm for a lot of the very old in this sample (although perhaps not Mayo Clinic people!) and less true for the younger ones. The authors should look for this and comment on it.

7. Similarly, marital and gender may interact in their effects in some cohorts (ie marriage is better for men than for women in the depression era cohorts). The authors should look at interaction terms for this.
8. There should be a flow diagram of the sample (and loss to follow-up through death, incomplete data, etc.)
9. There should be some note of the effect of drop-outs and deaths on the analyses of trajectories. Terminal drop and terminal decline can (possibly) change the trajectories. Analysis of this is complicated, and it may be best just to state that it is a limitation.

Minor Points

1. The FI should likely be called the Frailty Index, rather than the Rockwood Index.
2. I am not sure if the English should be UK or US spelling.
3. Some would quibble with the term SES. It comes directly from Weberian theory (and the authors don't consider other social groupings). It may be preferable to use the term Social Position (which is apparently less theory based) – see the series in the Int J Epi. This would depend upon the journal policy.
4. Similarly , the term gender may be considered since we are not dealing with the biological trait. Again, the journal policy may dictate using "sex".
5. A highish BMI (though maybe not very very high) is usually protective against mortality in nearly all large epidemiological studies in older old people. Here, however, they count a high BMI as a deficit. From Table 1, this appears to the case here as well. The authors should explain why they count a high BMI as a deficit, and not as an asset (at least in the BMI up to 27).
6. The measurement of alcohol misuse may be too specific and count only the very high misusers. There is little the authors can do about this, but they should acknowledge this as a limitation.
7. If education can be left as a continuous variable (its not clear how this is entered into the data set), that may be preferable. The effect of education is very large indeed, and it is likely a gradient (which has policy implications), so it would be important to know if it's a gradient or threshold effect. If it can't, then it should be acknowledged as a limitation.
8. Living with family may be confounding by indication. Those whom family are worried about may move in with family and then become frail. The authors should comment on this.

Overall, this is a solid addition to the literature. It is not too surprising, but it is important. There are few prospective studies from which one can derive frailty measures, and the issue of trajectories of frailty is very important for both clinicians and policy makers.

REVIEWER	Norrina Allen Feinberg School of Medicine Northwestern University Chicago, IL USA
REVIEW RETURNED	08-Mar-2016

GENERAL COMMENTS	Examining the patterns or trajectories of frailty in older age and their predictors is an important area of research. The authors apply a sophisticated modeling strategy to address this question and found importantly that patient demographics and lifestyle factors are related to frailty trajectories during the 60's and 70's. Interestingly, these factors were not related to frailty trajectories among the over 80 population. While it offers important contributions to the field, there are several concerns that should be addressed. 1) Why did the authors choose to break stratify ages into 3 discrete categories and construct trajectories in each of the three strata separately? What happened to people who had frailty measures across multiple age strata? If an individual was assigned to the high frailty trajectory in 60-69, how likely was it that they were in the high frailty trajectory in 70-79, etc? Across the entire age range, on average how many frailty measures did participants have? 2) Excluding individuals who had died seems to restrict the trajectories that would be identified and limits the generalizability of the sample (differently across the age strata). This should be addressed. 3) Throughout the manuscript the trajectories are referred to as the lowest and highest, this implies that there were more than 2. Please discuss low vs high. 4) In the abstract and discussion causal language is used (i.e. "...social and behavioral factors LEADING to increasing frailty..." given the observational nature of this study the authors should be more conservative with the descriptions of their findings. 5) More details on the calculated frailty index are needed. For example in the methods the BMI and ADLs were stated to have been collected at one institution. Does that mean this analyses were only done from individuals at one institution or were they not included in the frailty scores for individuals at other institutions? The same question applies to the determination of social and behavioral factors. Why was a cut-point of 0.2 used for the logistic regression analyses of frailty? Please justify. In addition, when the frailty measure is calculated does it only include ICD-9 codes that occurred during that year or does a condition continue from the time of the first code and is counted in all following years? 6) Adjustment was made for baseline frailty. How much does the distribution of baseline frailty overlap between the high and low trajectory classes? 7) The authors state in the methods that the model with the best fit for 60-69 year olds included 3 trajectory classes. How was model fit assessed? Why did the authors choose to collapse it into 2 trajectories?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Arnold Mitnitski
Institution and Country: Dalhousie University, Canada Competing Interests: 'None declared'

This is generally thoroughly done study. The authors rightly mentioned that most studied examining social and behavioral factors were cross-sectional and just a few studies considered longitudinal changes. Some of existing studies escaped attention of the authors (Mitnitski et al. Exp Ger 2013) although without considering social and behavioral factors.

We appreciate the positive comments from the reviewer and have added the suggested reference to the Introduction. 'We and others have previously shown that frailty trajectories are strongly associated with multiple adverse outcomes.[3, 8-10]'

One of the major finding is that these factors were stronger associated with frailty changes in younger group. No such association was found in the 80-89 year old group. This result might undermine (in my opinion) their conclusion that social and behavioral factors leading to increasing frailty may inform interventions for individuals at risk of worsening frailty, although the authors discussed this in the paragraph before the Limitations and Strengths section.

We agree and while we discussed this in the paragraph before the 'Limitations and strengths,' we failed to mention this in the conclusion. Thus, we have made edits to the conclusion in the abstract as well as the concluding sentence of the manuscript as follows. 'Social and behavioral factors associated with increasing frailty may offer a way to target interventions for aging individuals at risk of worsening frailty, specifically when targeted at younger individuals.'

I am not sure that the reference [32] in respect to '…can be constructed using deficits available in the electronic medical record' is correct (I have not found such a statement in [32]) although the frailty index indeed had been created using the electronic medical records by J Young and A Clegg and presented at the British Geriatric Society meeting(s)).

We carefully reviewed this reference and while the authors state that 'a frailty index can be constructed using information that is readily available in most health surveys' and 'the process allows operationalization of the frailty index to be carried out in other datasets,' we did not find a statement specifically citing the use of the electronic medical record to create a frailty index. Thus, we have removed this citation and have instead cited a recent publication by Dr. Clegg:
Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E, Mohammed MA, Parry J, Marshall T. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age Ageing. 2016

I am not sure that I understand the last 3 sentences (lines 42-53).

We have provided more details in an attempt to clarify this in the methods. Specifically, while 3 trajectories was the best fit for the 60-69 year olds, the proportional odds assumption was not met when we fit an ordinal logistic regression. We assessed the fit for 2 trajectories and found that the majority of the patients also fit 2 trajectories with high probability. Thus, we chose to model 2 trajectories in the 60-69 year olds to be able to model the association between trajectory groups and social and behavioral factors and for consistency with the older age groups. The following changes have been made in the text. 'For the 60-69 year olds, the proportional odds assumption to fit an ordinal logistic regression model was not met; therefore model fit was assessed with 2 trajectories. The proportion of patients in each trajectory that had a predicted probability of >80% of being in that trajectory was 98% for the low trajectory and 85% for the high trajectory when 2 trajectories were

used. Therefore, while the best fit was for the model with 3 trajectories, the majority of patients fit in the 2 trajectories with high probability. Thus, for consistency, we remodeled the frailty trajectories in the 60-69 year olds using only 2 trajectories.'

Figure is identical to just published in the J Am Geriatr Soci 2016;64:

Yes, the figures are the same for the 70-79 and 80-89 year olds. However, the figure for the 60-69 year olds is different because we have modeled 2 trajectories for this manuscript instead of 3 trajectories as in the previous paper. We prefer to include the figures in this manuscript for clarity despite the fact that the figure is the same in the oldest 2 age groups. We would anticipate that having the figures included in the manuscript would make it easier for the readers to understand our methods and follow the results. However, we are willing to reference this previous paper and exclude the figure at the Editor's discretion.

Ethics board review statement is missing.

This can be found in last sentence in the 'Study population' subsection of the Methods:
'This study was approved by the Mayo Clinic and Olmsted Medical Center Institutional Review Boards.'

Reviewer: 2
Reviewer Name: Phil St John
Institution and Country: University of Manitoba, Canada Competing Interests: None declared

There are relatively few prospective cohort studies with the data required to determine frailty trajectories, and fewer still with social and behavioural factors, so this study is a useful addition to the literature.

Briefly, the authors constructed a Frailty Index (FI) from administrative data from a well known population based health care registry, and show that social factors are associated with frailty in prospective data analysis. In particular, education and concerns about alcohol misuse strongly predict worsening frailty. The paper is well written, easy to follow, and is fairly important. While not too surprising, the findings are important clinically and for policy.

Major Points
1. I do not have Figure 1.

We apologize for any issues in uploading the figure into the system. We hope that you are now able to view the figure, which has been relabeled as Figure 2 and has been uploaded as a separate file.

2. These analyses are based on administrative data. These are often not as accurate or carefully gathered as primary data. The authors should acknowledge this as a limitation.

We added the following sentence in the limitations section. 'However, we relied solely on administrative data to construct the frailty index and possible inaccuracies with this data may have affected our results.'

3. There is considerable debate (and indeed disagreement) regarding the conceptual model and definition of frailty. The "Frailty as a Phenotype " model is quite different, and may have different trajectories. The FI was chosen (I suspect) because the data were available. Ideally, one would have both measures of frailty. The authors acknowledge this a little, but should comment a bit more on their

choice of a measure of frailty. This is in no wise a fatal flaw, simply something which requires explanation.

We appreciate the concerns that the frailty index is not the same as the frailty phenotype and that we may have gotten different trajectories if we were able to measure the frailty phenotype. However, with over 12,000 patients in our study, it would be impossible to implement a measure of frailty that requires in-person physical assessments. In a previous study from our group in 223 patients with heart failure from Olmsted County, MN, we found correlation between the frailty index and the frailty phenotype (McNallan, et al. Am Heart J, 2013). The frailty index increased across the phenotype groups (mean frailty index was 0.17, 0.25, and 0.37 in the not frail, intermediate frail, and frail phenotype groups; $p<0.001$). Importantly, both measures of frailty predicted death equally well (C-statistics: 0.700 for frailty index and 0.687 for frailty phenotype). Thus, a distinct advantage of using the frailty index in our current study is our ability to capture changes over time using data available from the electronic medical record on over 12,000 individuals.

We have made additional edits to the limitations section of the discussion as follows. 'Potential limitations of our study include the choice of our frailty metric. There are different ways to measure frailty and our results may have differed if we used a different definition of frailty. The biologic phenotype of frailty,[15] which incorporates physical assessments, may be considered the gold standard in measuring frailty. However, we relied solely on administrative data to construct the frailty index and possible inaccuracies with this data may have affected our results.'

4. There are quite a number of ways to analyse these data (none particularly better than the other), and it is very complicated. The paper should likely be sent for statistical review. An appendix on statistical techniques and whether the modeling assumptions were met would be a good idea. As well, the authors should likely do some sensitivity analyses (GEE and multilevel models, for instance) to see if the results are the same. If they are (which they likely will be), then it strengthens the argument.

We agree with the reviewer that there are a number of ways to analyze this longitudinal data. The goal with our analytic approach was to group people (identify clusters) who follow a similar change over time in frailty instead of summarizing these longitudinal changes with the average change in all patients. In particular, we had previously identified clusters of frailty over time and wanted to expand on this research to examine associations between social and behavioral factors known to be associated with health outcomes and frailty. There are multiple ways to identify clusters of frailty over time, with Proc Traj (SAS Institute, Cary, NC) and K means cluster modeling (KmL, R, version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria)) being the two most common cited in the literature. The advantage of using a K means cluster modeling approach is that it is non-parametric and does not require any assumptions regarding the parameterization within the clusters or any assumptions regarding the shape of the trajectories. We have previously compared the use of Proc Traj and KmL and found that they identified similar trajectories. Others, such as Hoogendijk, EO, et. al. (Ann Epidemiol. 2014 Jul;24(7):538-44.e2) have previously assessed frailty and education level using 2-level longitudinal mixed model regression. They assessed adults ages 65 and older over a 13 year period and found educational differences in frailty. Frailty rates were higher for those with lower education levels and the frailty prevalence increased over time. They did not find an association between the rate of increase in frailty prevalence and education level. It is however, difficult to directly compare the mixed model results to our analysis. The mixed models can be used to compare frailty across levels of education (or other social factors) and to compare the rate of change in frailty over time across levels of education. Our analysis first identifies different clusters of frailty trajectories and then compares differences in education among those clusters. We modified our description of the KmL method in the 'Statistical analysis' section. 'K means cluster modeling for longitudinal data (KmL, R, version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria)), was used to identify distinct, homogenous clusters of frailty index trajectories within each age group.[42] KmL is a non-

parametric hill-climbing algorithm, and does not impose assumptions regarding the parameterization within the clusters or the shape of the trajectories.'

5. The criteria for an item to be included in the FI (common enough to saturate a model, adverse, stable over time etc) also include age. So its difficult to comment on age-stratified analyses, since its really looking at the same predictor variable twice. It would be reasonable to stratify disability, multimobidity, and so on, but I am not sure what it means to stratify frailty on age. Of course, the question is relevant and interesting. Since its only one criterion, it may not have a large influence, and I would be inclined to simply ask for a statistical review of this issue. (It may be more reasonable to drop this as a criterion for inclusion in the FI).

As expected, a higher baseline frailty index was observed with older age. In addition, as presented in our published manuscript (J Am Geriatr Soc. 2016 Feb;64(2):285-92), the younger individuals have a lower prevalence of most deficits compared to older patients and comorbidities played a larger role in frailty in the younger (60-69 and 70-79 years) age groups, whereas ADLs played a larger role in the oldest (80-89 years) age group. Furthermore, the risk of death, in particular, is highly influenced by age. Also, as can be seen in Table 2 of the current manuscript, there are also differences in the social factors across age, with lower education and lower proportion of patients who were married and living with a spouse with older age. Thus, because both the frailty index and social factors were influenced by age, we stratified our analyses by age. Because of our large sample size, we are powered to look at stratified analyses, which we believe is an advantage of our study. Furthermore, because comorbidities and ADLs are used to define the frailty index, we are not able to stratify our results on disability or multimorbidity. We made edits to the 'Statistical analysis' subsection of the Methods to further clarify our reasons to stratify the analyses by age. 'Patients with at least 3 years of frailty measures were retained for analysis (n=12,270). The frailty index and prevalence of social factors differed across age. Thus, we stratified our cohort into 3 groups based on age in 2005: 60-69 years, 70-79 years, and 80-89 years.'

6. In most countries, there has been a strong cohort effect of education. So usually education and age are highly correlated. This is particularly true on the Great Plains, where an incomplete education in one room schools was the norm for a lot of the very old in this sample (although perhaps not Mayo Clinic people!) and less true for the younger ones. The authors should look for this and comment on it.

As can be seen in table 2, there is a correlation with age and education. The younger individuals were more highly educated and the oldest individuals had a higher proportion of less than a high school education. In addition, lower proportions of married individuals and living with a spouse/domestic partner was observed with increasing age. In addition, the baseline frailty index increased with increasing age. For these reasons, we stratified our analysis by age at baseline. We have added more details in the 'Statistical analysis' subsection of the Methods to describe why we chose to stratify on age and to clarify that the stratification was based on age at baseline (2005). 'Patients with at least 3 years of frailty measures were retained for analysis (n=12,270). The frailty index and prevalence of social factors differed across age. Thus, we stratified our cohort into 3 groups based on age in 2005: 60-69 years, 70-79 years, and 80-89 years. The baseline age defined the age groups and the same groupings were used across follow-up although patients could have aged into another age group during follow-up. K means cluster modeling for longitudinal data (KmL, R, version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria)), was used to identify distinct, homogenous clusters of frailty index trajectories within each age group.[42]'

7. Similarly, marital and gender may interact in their effects in some cohorts (ie marriage is better for men than for women in the depression era cohorts). The authors should look at interaction terms for this.

As mentioned at the end of the 'Statistical analysis' section of the methods, we did test interactions between sex and each social and behavioral factor. We did not find any significant interactions, and thus, had not presented any stratified results by sex. Below are the unadjusted odds ratios for being in the high frailty trajectory for non-married vs. married/committed marital status in men and women separately, along with a corresponding p-value testing the sex*marital status interaction.

60-69 year olds 70-79 year olds 80-89 year olds
Men, OR (95% CI) 1.90 (1.49-2.43) 1.83 (1.40-2.39) 1.31 (0.91-1.88)
Women, OR (95% CI) 2.01 (1.71-2.37) 1.78 (1.50-2.11) 1.69 (1.31-2.17)
P-value for interaction 0.712 0.850 0.258

8. There should be a flow diagram of the sample (and loss to follow-up through death, incomplete data, etc.)

Key numbers on the breakdown of our cohort are provided in the first paragraph of the results. In addition, we have now created a figure (Figure 1) which describes the number of patients excluded including a breakdown of the number of those who died within 3 years and those with >3 years of follow-up. We have also provided a distribution of the number of years of frailty index for those excluded and those included in the analysis. The median number of years of frailty for those included in our analysis was 5 and this information has also been added to the first paragraph of the Results section. Finally, among those included in the analysis, we have provided information on the proportion that died at some point during the follow-up. 'Among the 16,443 residents of Olmsted County, MN aged 60-89 in 2005, 12,270 (74.6%) had at least 3 years of frailty measures between 2005 and 2012 and were retained for analysis (Figure 1). Of those retained, the median number of years of frailty measures was 5. Of those excluded, 17% had died within 3 years of follow-up. Of the 12,270 patients retained, 44.5% were male, 49.3% were aged 60-69 and 15.6% were aged 80-89 at baseline, and 18% died during follow-up.'

9. There should be some note of the effect of drop-outs and deaths on the analyses of trajectories. Terminal drop and terminal decline can (possibly) change the trajectories. Analysis of this is complicated, and it may be best just to state that it is a limitation.

We agree with the reviewer that terminal drop or terminal decline could affect a patient's frailty trajectory and it would be difficult to distinguish between changes due to terminal decline and changes related to the aging process. There are a couple of aspects of our analysis that should lessen the effect of rapid changes in frailty before death in assessing frailty trajectories over long-term follow-up. First, as noted in the methods, we only used one frailty index for each year for each person in our trajectory analyses. With this approach, we would expect that terminal decline would primarily be seen only in the last frailty index for a given person. Along with this we required at least 3 years of frailty index to be included in the frailty trajectory analyses. This should also help to reflect the long-term changes related to the aging process. We noted this in as a limitation. 'Finally, while we clustered individuals by their long-term frailty trajectories, it is difficult to distinguish between rapid changes in frailty before death and changes related to the aging process.'

Minor Points

1. The FI should likely be called the Frailty Index, rather than the Rockwood Index.

We have made this correction and now refer to the 'frailty index' throughout the manuscript.

2. I am not sure if the English should be UK or US spelling.

We are happy to make changes as necessary and would appreciate the Editor providing information on such changes if appropriate.

3. Some would quibble with the term SES. It comes directly from Weberian theory (and the authors don't consider other social groupings). It may be preferable to use the term Social Position (which is apparently less theory based) – see the series in the Int J Epi. This would depend upon the journal policy.

We have replaced the term 'socioeconomic status' with 'social position' in the discussion.

4. Similarly, the term gender may be considered since we are not dealing with the biological trait. Again, the journal policy may dictate using "sex".

Our strong preference is to use the term 'sex' since we are actually referring to biological sex as opposed to gender behaviors, roles, identity, etc.

5. A highish BMI (though maybe not very very high) is usually protective against mortality in nearly all large epidemiological studies in older old people. Here, however, they count a high BMI as a deficit. From Table 1, this appears to the case here as well. The authors should explain why they count a high BMI as a deficit, and not as an asset (at least in the BMI up to 27).

While we agree that a somewhat elevated BMI may be associated with better outcomes in elderly patients, we chose to be consistent with previous publications that used the frailty index (Searle et al, BMJ Geriatrics 2008; McNallan et al, JACC Heart Fail 2013), and thus assigned 1 point for those with a BMI <18.5 and BMI ≥30.

6. The measurement of alcohol misuse may be too specific and count only the very high misusers. There is little the authors can do about this, but they should acknowledge this as a limitation.

We agree that the way the questions about alcohol are written, we may only be capturing the highest misusers. We have added this as a limitation to our study, as follows. 'In addition, our questions about alcohol consumption may have only identified individuals with excessive consumption and alcoholism. We did not have information to quantify the amount of drinking in all individuals to identify other patterns of drinking.'

7. If education can be left as a continuous variable (its not clear how this is entered into the data set), that may be preferable. The effect of education is very large indeed, and it is likely a gradient (which has policy implications), so it would be important to know if it's a gradient or threshold effect. If it can't, then it should be acknowledged as a limitation.

Our question on education includes categorical responses and we do not have a continuous variable with number of years of education in order to treat education as a continuous variable in the analyses. Thus, we are unable to assess whether there is a gradient or threshold effect, which we have added to our limitations. 'Third, there were a limited number of social factors available from the questionnaire; some potentially important social factors (for example, income) were not available and others may not have been optimally measured. For example, education was assessed as a categorical variable which limited our ability to determine if a gradient or threshold effect was apparent for the association of years of education with frailty trajectories.'

8. Living with family may be confounding by indication. Those whom family are worried about may move in with family and then become frail. The authors should comment on this.

We agree and it is difficult to discern whether a patient moved in with family because they were becoming frail or if they became frail after moving in with family. Thus, we have thus modified our discussion about this finding as follows. 'Furthermore, our study is the first to report that living alone or with family (compared to living with a spouse) is associated with longitudinal changes in frailty, with an increased risk of being in the high frailty trajectory in all age groups. However, the associations persisted only among the 70-79 year olds once we adjusted for baseline frailty. It should be noted, however, that living with family may be a marker of increasing frailty and patients may move in with family because they are becoming frail.'

Overall, this is a solid addition to the literature. It is not too surprising, but it is important. There are few prospective studies from which one can derive frailty measures, and the issue of trajectories of frailty is very important for both clinicians and policy makers.

We appreciate the positive comments about our manuscript.

Reviewer: 3
Reviewer Name: Norrina Allen
Institution and Country: Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
Competing Interests: None declared

Examining the patterns or trajectories of frailty in older age and their predictors is an important area of research. The authors apply a sophisticated modeling strategy to address this question and found importantly that patient demographics and lifestyle factors are related to frailty trajectories during the 60's and 70's. Interestingly, these factors were not related to frailty trajectories among the over 80 population. While it offers important contributions to the field, there are several concerns that should be addressed.

1) Why did the authors choose to break stratify ages into 3 discrete categories and construct trajectories in each of the three strata separately? What happened to people who had frailty measures across multiple age strata? If an individual was assigned to the high frailty trajectory in 60-69, how likely was it that they were in the high frailty trajectory in 70-79, etc? Across the entire age range, on average how many frailty measures did participants have?

The mean frailty index at baseline increased with age and the distribution of social factors also differed by age (lower education and lower proportion of patients who were married and living with a spouse was observed with increasing age). Thus, we decided to stratify our results on age. We created the strata based on age at baseline (in 2005) and individuals did not move into different groups as they aged. We have added more details in the 'Statistical analysis' subsection of the Methods to describe why we chose to stratify on age and to clarify that the stratification was based on age at baseline. 'Patients with at least 3 years of frailty measures were retained for analysis (n=12,270). The frailty index and prevalence of social factors differed across age. Thus, we stratified our cohort into 3 groups based on age in 2005: 60-69 years, 70-79 years, and 80-89 years. The baseline age defined the age groups and the same groupings were used across follow-up although patients could have aged into another age group during follow-up. K means cluster modeling for longitudinal data (KmL, R, version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria)), was used to identify distinct, homogenous clusters of frailty index trajectories within each age group.[42]'

Across the entire age range, the mean number of frailty measures was 5.3 and the median was 5. This has been added to the first paragraph of the Results section. 'Among the 16,443 residents of Olmsted County, MN aged 60-89 in 2005, 12,270 (74.6%) had at least 3 years of frailty measures

between 2005 and 2012 and were retained for analysis. Of those retained, the median number of years of frailty measures was 5.' In addition, we have also added a new Figure 1 which provides a distribution of the number of years of frailty index for those excluded and those included in the analysis.

2) Excluding individuals who had died seems to restrict the trajectories that would be identified and limits the generalizability of the sample (differently across the age strata). This should be addressed.

We apologize for the confusion as we did not exclude all individuals who died. We did limit the analysis to individuals who had at least 3 years of frailty index measures. This excluded 4,173 individuals, of whom 703 (17%) had died. Of the 12,270 who had at least 3 years of frailty indexes and were included in the analysis, 2,168 (18%) who died during follow-up were not excluded. We have added a new Figure 1 which provides more details on the patients excluded and we have also edited the text of the first paragraph of the Results section to clarify that 18% of the patients included in the analysis died during follow-up. 'Of the 12,270 patients retained, 44.5% were male, 49.3% were aged 60-69 and 15.6% were aged 80-89 at baseline, and 18% died during follow-up.'

3) Throughout the manuscript the trajectories are referred to as the lowest and highest, this implies that there were more than 2. Please discuss low vs high.

We have replaced 'lowest' and 'highest' with 'low' and 'high' throughout the manuscript.

4) In the abstract and discussion causal language is used (i.e. "...social and behavioral factors LEADING to increasing frailty..." given the observational nature of this study the authors should be more conservative with the descriptions of their findings.

In the abstract and discussion, we have toned down the language and replaced 'leading to increasing frailty' with 'associated with increasing frailty.' Likewise, we have also reworded the last sentence in the 'Limitations and strengths' subsection of the discussion replacing 'influence longitudinal changes in frailty' with 'are associated with longitudinal changes in frailty.'

5) More details on the calculated frailty index are needed. For example in the methods the BMI and ADLs were stated to have been collected at one institution. Does that mean this analyses were only done from individuals at one institution or were they not included in the frailty scores for individuals at other institutions? The same question applies to the determination of social and behavioral factors. Why was a cut-point of 0.2 used for the logistic regression analyses of frailty? Please justify. In addition, when the frailty measure is calculated does it only include ICD-9 codes that occurred during that year or does a condition continue from the time of the first code and is counted in all following years?

Patients residing in Olmsted County, MN may receive care at more than 1 institution in the county. Utilizing the resources of the Rochester Epidemiology Project, we are able to identify all encounters of care occurring at different institutions for a given individual, and some information including ICD-9 codes are available from all sources of care. However, BMI and ADLs were only available from one institution. For BMI data, we imputed missing BMI values between the first and last available BMI using linear interpolation. For the comorbidities, once a diagnosis was made, we carried this forward for all subsequent years. In addition, if a patient was missing 3 or fewer variables in the frailty index, we still calculated the frailty index and adjusted the denominator accordingly. If more than 3 variables were missing, a frailty index for that year was not calculated. These details were included in our previously published manuscript (J Am Geriatr Soc. 2016 Feb;64(2):285-92), but we failed to specify these details in this manuscript. Thus we have edited our Methods to provide more details on the calculation of the frailty index. 'The index was calculated only if 3 or fewer items were missing,

adjusting the denominator accordingly. Repeated measures of frailty index were calculated for each year through 2012 (one index per year). Linear interpolation was used to impute missing BMI values between the first and last available BMI. Once a comorbidity was identified, this was carried forward for all subsequent frailty index measures. More details on the rules for calculating the frailty index have been reported previously.[9]'

The cutpoint of 0.20 to distinguish between high and low baseline frailty was chosen for 2 reasons. First, although the frailty index is generally not categorized, this cutpoint is recognized as approaching a frail state (BMC Geriatr. 2008 Sep 30;8:24). Prior comparisons of the frailty index to the frailty phenotype reported that those categorized as pre-frail phenotype had a mean frailty index of 0.25 in patients with heart failure (Am Heart J. 2013 Oct;166(4):768-74), 0.25 in elderly patients admitted to a geriatric day hospital (J Am Med Dir Assoc. 2015 Oct 1;16(10):855-9), and 0.30 in individuals over 70 years of age (J Gerontol A Biol Sci Med Sci. 2007 Jul;62(7):738-43). Second, in our dataset, we looked at the distribution of frailty index at baseline overall and this value corresponded to the 75th percentile (75th percentile=0.19). We have made edits to the 'Statistical analysis' subsection of the Methods as follows. 'Logistic regression was used to determine the associations of social and behavioral factors with baseline frailty (defining high baseline frailty as a frailty index of 0.20 or higher, which corresponds to the 75th percentile of the distribution of the frailty index) and frailty trajectories (modeling the odds of being in the high frailty trajectory).'

6) Adjustment was made for baseline frailty. How much does the distribution of baseline frailty overlap between the high and low trajectory classes?

As can be seen in the figures and noted in our previous paper (J Am Geriatr Soc. 2016 Feb;64(2):285-92), baseline frailty increases with age and is higher for those in the upper trajectory of each age group. We have provided below the median, IQR, and minimum and maximum frailty index at baseline for those in the low and high trajectories in each age group. There is significant separation, with some overlap in baseline frailty between the low and high frailty trajectories in each age group. Importantly, we show in Table 4 that social and behavioral factors are associated with frailty trajectories beyond what can be explained by baseline frailty levels, specifically for those less than 80 years of age.

60-69 year olds 70-79 year olds 80-89 year olds
Baseline frailty index Low trajectory High trajectory Low trajectory High trajectory Low trajectory High trajectory
Median 0.08 0.19 0.11 0.23 0.14 0.27
Q1, Q3 0.05, 0.13 0.16, 0.25 0.08, 0.16 0.19, 0.3 0.09, 0.17 0.22, 0.34
Min, max 0, 0.39 0, 0.68 0, 0.41 0, 0.8 0, 0.45 0.08, 0.75

7) The authors state in the methods that the model with the best fit for 60-69 year olds included 3 trajectory classes. How was model fit assessed? Why did the authors choose to collapse it into 2 trajectories?

The model fit was assessed using the Calinski and Harabatz criterion and the optimum number of clusters for the 60-69 year olds was 3. However, when we fit an ordinal logistic regression model to determine the associations of social and behavioral factors with frailty trajectories (using the 3 trajectories), the proportional odds assumption was not met for this model. We then assessed the model fit with 2 trajectories and found the fit to be acceptable. Thus, for consistency purposes, we decided to model 2 trajectories in the 60-69 year olds. The following changes were made to the 'Statistical analysis' subsection of the Methods for clarification. 'The optimum number of clusters in each age group was determined using the Calinksi and Harabatz criterion.[42] The optimum number of clusters was 3 for the 60-69 year olds and 2 for both the 70-79 and 80-89 years. For the 60-69 year

olds, the proportional odds assumption to fit an ordinal logistic regression model was not met; therefore model fit was assessed with 2 trajectories. The proportion of patients in each trajectory that had a predicted probability of >80% of being in that trajectory was 98% for the low trajectory and 85% for the high trajectory when 2 trajectories were used. Therefore, while the best fit was for the model with 3 trajectories, the majority of patients fit in the 2 trajectories with high probability. Thus, for consistency, we remodeled the frailty trajectories in the 60-69 year olds using only 2 trajectories.'

VERSION 2 – REVIEW

REVIEWER	Arnold Mitnitski Dalhousie University, Canada
REVIEW RETURNED	09-Apr-2016

GENERAL COMMENTS	The authors well responded to my comments/concerns but I found one point that appeared in the revised version of the manuscript in response to the other reviewer's major point #3. "The biologic phenotype of frailty,[15] which incorporates physical assessments, may be considered the gold standard in measuring frailty." I don't know where the authors found that the frailty phenotype is (1) biologic and (2) is the gold standard in measuring frailty. As for being biologic phenomenon (as sometimes claimed) is not more than an opinion of some authors – outside of its definition is no more or less "biological" than the other measures of frailty. The most importantly, it is by no means "the gold standard" – to my knowledge that has not been claimed even by the authors of that approach. I would encourage the authors to make changes in that paragraph- at the same time, I am aware that the authors bear their responsibility for the content they decide to present and I do not want to use the review process as a battleground with the other frailty camps. That is only a suggestion to make the manuscript more coherent. For that matter, I would like to attract the authors' attention to the papers: Theou O, Brothers TD, Peña FG, Mitnitski A, Rockwood K. Identifying common characteristics of frailty across seven scales. J Am Geriatr Soc. 2014 May;62(5):901-6. Theou O, Walston J, Rockwood K. Operationalizing Frailty Using the Frailty Phenotype and Deficit Accumulation Approaches. Interdiscip Top Gerontol Geriatr. 2015;41:66-73. Theou O, Rockwood K. Comparison and Clinical Applications of the Frailty Phenotype and Frailty Index Approaches. Interdiscip Top Gerontol Geriatr. 2015;41:74-84. In spite these comments I vote to accept the manuscript.

REVIEWER	Phil St John University of Manitoba

REVIEW RETURNED	22-Apr-2016

GENERAL COMMENTS	The authors have addressed the issues.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Arnold Mitnitski
Institution and Country: Dalhousie University, Canada Competing Interests: None declared

The authors well responded to my comments/concerns but I found one point that appeared in the revised version of the manuscript in response to the other reviewer's major point #3.

“The biologic phenotype of frailty,[15] which incorporates physical assessments, may be considered the gold standard in measuring frailty.” I don’t know where the authors found that the frailty phenotype is (1) biologic and (2) is the gold standard in measuring frailty.

As for being biologic phenomenon (as sometimes claimed) is not more than an opinion of some authors – outside of its definition is no more or less “biological” than the other measures of frailty. The most importantly, it is by no means "the gold standard" – to my knowledge that has not been claimed even by the authors of that approach.

I would encourage the authors to make changes in that paragraph- at the same time, I am aware that the authors bear their responsibility for the content they decide to present and I do not want to use the review process as a battleground with the other frailty camps. That is only a suggestion to make the manuscript more coherent.

We appreciate the opportunity to clarify this point in our discussion and have made the following changes to the manuscript. ‘Potential limitations of our study include the choice of our frailty metric. There are different ways to measure frailty and our results may have differed if we used a different definition of frailty. For example, the frailty phenotype[15] incorporates physical assessments, but with our large sample size, we were unable to implement a measure of frailty that requires in-person physical assessments. Thus, we relied solely on administrative data to construct the frailty index and possible inaccuracies with this data may have affected our results.’