PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Symptom burden in multimorbidity. A population-based combined questionnaire and registry study from Denmark
AUTHORS	Willadsen, Tora; Siersma, Volkert; Nicolaisdottir, Dagny; Jarbol, Dorte; Guassora, Ann; Reventlow, Susanne; Køster-Rasmussen, Rasmus

VERSION 1 – REVIEW

REVIEWER	Kasey Boehmer Mayo Clinic, USA
REVIEW RETURNED	04-Aug-2020

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript regarding symptom burden in patients with and without multimorbidity. Overall, the manuscript appears to oversimplify the work that was conducted, given the large amount of data that was collected through the lengthy questionnaire (with good response rate). The data could be more richly explored to make a greater contribution to the field. The article also misses some key points in the discussion of their findings, and has some methodological considerations that should be addressed in the manuscript prior to publication. Finally, I found the figures to be somewhat confusing and believe they need work prior to publication to better orient readers. Point-by-point comments are suggested below: Introduction Page 1, Line 12: “Symptoms are central for how to rate own health” – this reads oddly. Should it be “Symptoms are central to how one rates their own health”? Methods Page 6, Line 16: “Information on diagnoses is retrieved” should be past tense “information on diagnoses was retrieved.” Page 6. Line 18: Another instance of needing to be past tense - ensure past tense is used throughout since study has occurred – there are quite a few instances of this throughout the paper, so I have only pointed out a couple. Should be “Participants were excluded.” Overall – given the length of the survey, I would expect there was missing data within varying questions and scales. How were missing data handled? It says in the checklist that this information is stated on page 6, but it is not. Results Page 7, line 36: the numbers, such as 47.425 should have a comma instead of period e.g 47,425. Page 7 and 8 results: I am having a hard time understanding – participants in this study included those that were considered healthy (no diagnoses) as well as those that were not (diagnosis or diagnoses). Yet, participants with no diagnoses still averaged a

	mean number of almost 5 symptoms in the previous 4 weeks? This is a bit perplexing to me as I would expect healthy individuals to have low to no symptoms in the previous 4 weeks. Did they have conditions that were simply not captured by this classification structure? I think with such high presence of symptoms of those with no diagnoses, the authors need to address what characteristics that these patients have that could explain them. Page 8 – Instead of “mental” being the label used, can we use “mental health” diagnoses? The “mental-other dx” structure does not strike me as very patient centered. Similarly, it would be more patient centered to say “persons living with” or “patients living with” where sometimes the language is of belonging to certain categories. Figures 2-4 – I understood the text that refers to the figures better than the figures themselves. I think a key for what “Syn” “Eff” and “n” mean as well as some additional explanation as part of the figure would be helpful. Page 9, line 43-48 – I think a summary patient demographics table is a necessary part of the main text. The supplemental files are extremely detailed, but in order to understand the sample, the reader needs a better set of demographic characteristics (typically table 1 of a manuscript such as this one). This also perhaps would help the reader understand the findings in regards to patients with and without diagnoses as well. Overall – what about clusters of >2 groups of diagnoses? Given the large sample size, I would expect these to exist and would like to know how things compared to the 1 and 2 groups. Overall – The questionnaire collects data on healthcare access regarding symptoms (e.g. GP contact), but this information is entirely missing from the manuscript. I think this should be reported on. Discussion Page 11, lines 20 – 26 – again, please use “mental health diagnosis” instead of “mental diagnosis” Page 11, line 36 – rather than stating that patients become problem patients, please change this to be more patient-centered. Specifically, this could be stated as gradually becoming perceived by their clinicians as problematic patients, rather than those with problematic human situations. Questionnaire – Why was information regarding number of sexual relationships gathered for women only and not men? This seems very problematic and also irrelevant to the research question as far as I can tell based on what is reported. Overall – The discussion section is lacking mention of treatment burden, as symptom burden and treatment burden often go hand in hand (see, Cumulative Complexity Model Shippee 2012), especially as patients with increasing symptoms have an increased need to seek care for relief. I think the discussion of these interplays could be strengthened.

REVIEWER	Daniel Solomon Brigham and Women's Hospital Harvard Medical School Boston MA USA
REVIEW RETURNED	02-Nov-2020

GENERAL COMMENTS	The question is interesting and the methods are well described. It appears that the results are valid.

	It is not clear to me that the findings are important for research or clinical care. Having more morbidities is bad. It is interesting that symptoms are supra-additive, but its not clear what the implications are. Should certain morbidities be targeted to reduce symptoms? Are symptoms a bad measure of disease implications? I would also wonder whether the 10 categories of morbidities are a good measure of multimorbidity. Many morbidities within the same category are very different and the morbidity categories do not take into account different severities.

REVIEWER	Sarah Monsell University of Washington, USA
REVIEW RETURNED	04-Nov-2020

GENERAL COMMENTS	Abstract: The last sentence of the Results seems to convey the same information as the Conclusion. Methods: How did those who responded to the survey differ (or not) from those who did not respond to the survey in terms of patient demographics? It seems that the authors should have the data to perform this analysis. Methods: Please expand on the reason for excluding gender-specific symptoms. Was their prevalence relatively high or low? If substantial, separate models could be run for men and women given the large sample size, at least as a sensitivity analysis. Methods: Am I correct in reading that interaction effects between two specific diagnosis groups were evaluated? How frequent was the presence of 3 or more diagnosis groups for an individual? I think you may need to add the caveat to your conclusion that this "beyond additive" effect applies only to those with two diagnosis groups, and that the effect of having a combination of 3 or more diagnoses was not evaluated. Methods: I think Poisson regression would better fit the outcome of number of symptoms. Similarly, ordinal logistic regression is likely a better fit for the other two outcomes that have 4 possible levels. Figures 2, 3, and 4: These are really nice figures. My only concern is that we can't see confidence intervals. Please spell out "Syn" and "Eff." It would also be helpful to describe the interpretation of the effect. For example, an effect size of 4 translates to an average increase of 4 symptoms for participants with those two diagnoses compared to those with no diagnoses. This would be especially helpful given the different scales in each figure (and potentially link function as referenced above). Results: We need to see confidence intervals from the models. Results: Did the authors consider running models were the number of diagnosis groups (regardless of which specific groups), adjusted for the same covariates, was regressed on the symptom outcomes? I think this type of model would be highly informative. It would allow us to see if the relationship between the number of diagnosis groups and outcome in groups larger than two. After seeing the results for two diagnosis groups, I am quite interested in knowing if this effect holds as the number of diagnosis groups increases.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Kasey Boehmer
Institution and Country: Mayo Clinic, USA
Please state any competing interests or state 'None declared': None declared

Comments to the Author
Thank you for the opportunity to review this manuscript regarding symptom burden in patients with and without multimorbidity. Overall, the manuscript appears to oversimplify the work that was conducted, given the large amount of data that was collected through the lengthy questionnaire (with good response rate). The data could be more richly explored to make a greater contribution to the field. The article also misses some key points in the discussion of their findings, and has some methodological considerations that should be addressed in the manuscript prior to publication. Finally, I found the figures to be somewhat confusing and believe they need work prior to publication to better orient readers. Point-by-point comments are suggested below:

Introduction
1. Page 1, Line 12: "Symptoms are central for how to rate own health" – this reads oddly. Should it be "Symptoms are central to how one rates their own health"?

Yes it reads oddly, we have changed it to " Symptoms play a major part in how people self-rate their health". Thanks.

Methods
2. Page 6, Line 16: "Information on diagnoses is retrieved" should be past tense "information on diagnoses was retrieved."

We changed the sentence to past tense.

3. Page 6. Line 18: Another instance of needing to be past tense - ensure past tense is used throughout since study has occurred – there are quite a few instances of this throughout the paper, so I have only pointed out a couple. Should be "Participants were excluded."

Thanks for pointing this out. We have carefully revised the manuscript for typographical and grammatical errors, and we have rephrased parts of it to improve readability.

4. Overall – given the length of the survey, I would expect there was missing data within varying questions and scales. How were missing data handled? It says in the checklist that this information is stated on page 6, but it is not.

We have added this sentence on to the decription of the questionnaire in the methods section: The questionnaire was electronic, and designed so that it was not possible to skip items, and therefore there were no missing values for those who finished the questionnaire.

Results
5. Page 7, line 36: the numbers, such as 47.425 should have a comma instead of period e.g 47,425.

OK. All large numbers have been corrected to include commas instead of periods.

6. Page 7 and 8 results: I am having a hard time understanding – participants in this study included those that were considered healthy (no diagnoses) as well as those that were not (diagnosis or diagnoses). Yet, participants with no diagnoses still averaged a mean number of almost 5 symptoms in the previous 4 weeks? This is a bit perplexing to me as I would expect healthy individuals to have low to no symptoms in the previous 4 weeks. Did they have conditions that were simply not captured by this classification structure? I think with such high presence of symptoms of those with no diagnoses, the authors need to address what characteristics that these patients have that could explain them.

The questionnaire-study was population based and the persons were retrieved from the Danish Civil Registration System completely at random. We have no reason to believe that non-respondents had fewer diagnoses or symptoms than the respondents did. Please see our revised section about non-respondents below. As you point out, there is a high frequency of symptoms among participants not having a diagnosis from one of the ten diagnosis groups, and the frequency is not much lower than in participants with a morbidity or multimorbidity. We have included the result in the revised abstract. The explanation for this finding is likely

1) *That symptoms are prevalent in the background population (Elliott, 2011, McAteer, 2011, Hollnagel 1985) and in general practice (Elnegaard, 2015).*
2) *The national health registers we have used includes only diagnoses from the secondary sector*.
3) *The ten diagnosis groups we use to define multimorbidity is not an exhaustive list of all possible diseases (no definition of multimorbidity is). Therefore, a person in our study without a diagnosis from any of the ten diagnosis groups, could still have another disease e.g. allergy or pneumonia are not included in our definition. Other definitions of multimorbidity often use single count of diseases and include diagnoses like hypercholesterolemia. All diagnoses included in our definition are chronic diseases, and therefore we expected to see larger differences with our definition in comparison to other studies using single count of diseases.*

We elaborate more on this (marked with yellow) in the discussion.

7. Page 8 – Instead of “mental” being the label used, can we use “mental health” diagnoses? The “mental-other dx” structure does not strike me as very patient centered. Similarly, it would be more patient centered to say “persons living with” or “patients living with” where sometimes the language is of belonging to certain categories.

Thank you for making this central point. We have earlier published a study using the same definition and terminology as we use in this study, but where we included the entire Danish population and studied the relation between multimorbidity and mortality. However, in this study, symptom burden was the outcome of interest which is a rather subjective and patient focused area. Therefore, we agree that the diagnosis group can be expressed in a better way. We have changed the wording to mental health throughout the manuscript.

8. Figures 2-4 – I understood the text that refers to the figures better than the figures themselves. I think a key for what “Syn” “Eff” and “n” mean as well as some additional explanation as part of the figure would be helpful.
We have added the following to the figure legend.

Figure 2. *All (45) combinations of multimorbidity (two diagnosis groups) and the association with number of symptoms. Syn=synergy, the excess number of symptoms for multimorbidity, compared to the sum of symptoms from two people having the two morbidity domains individually. Eff=effect, the number of symptoms for the multimorbidity combination. N=number, the number of people with the multimorbidity combination. Effects with a p-value <0.05 are marked with *.*

Figure 3. *All (45) combinations of multimorbidity (two diagnosis groups) and the association with influence on daily activities. Syn=synergy, the excess interference with usual daily activities score (from the symptom with the highest interference score, ranging 1-4 with 4 indicating the highest burden on usual daily activities) for multimorbidity, compared to the sum of the interference score from two people with the two morbidity domains individually. Eff=effect, the interference score for the multimorbidity combination. N=number, the number of people with the multimorbidity combination. Effects with a p-value <0.05 are marked with *.*

Figure 4. *All (45) combinations of multimorbidity (two diagnosis groups) and the association with concern about symptoms. Syn=synergy, the excess concern score (from the symptom with the highest concern score, ranging 1-4 with 4 indicating most concern) for multimorbidity, compared to the sum of the concern score from two people with the two morbidity domains individually. Eff=effect, the concern score for the multimorbidity combination. N=number, the number of people with the multimorbidity combination. Effects with a p-value <0.05 are marked with *.*

9. Page 9, line 43-48 – I think a summary patient demographics table is a necessary part of the main text. The supplemental files are extremely detailed, but in order to understand the sample, the reader needs a better set of demographic characteristics (typically table 1 of a manuscript such as this one). This also perhaps would help the reader understand the findings in regards to patients with and without diagnoses as well.

We have included the table in supplementary file 5 (descriptive overview of patient characteristics for the participants in The Danish Symptom Cohort) as a new table 1 in the main manuscript.

10. Overall – what about clusters of >2 groups of diagnoses? Given the large sample size, I would expect these to exist and would like to know how things compared to the 1 and 2 groups.

Yes, there are people in our sample that have diagnoses in more than two morbidities, and we have included numbers for their symptom burden to the results section. However, the focus of the present analysis is the symptom burden attributable to specific combinations morbidities.
While synergy can be defined for combinations of three groups of diagnoses, these combinations (120) number many more than combinations of two (45) and have much lower numbers of people in them (some of them do not occur at all); inference will be weak. More importantly, overview of the results will be harder to achieve, tables and figures will have a third dimension and thereby will be large and confusing.

11. Overall – The questionnaire collects data on healthcare access regarding symptoms (e.g. GP contact), but this information is entirely missing from the manuscript. I think this should be reported on.

We have added this to the methods section:

The Danish Symptom Cohort was established to investigate symptom experience in the general population and healthcare-seeking in relation to general practice. According to the random sample selected from the Danish population, the cohort included both healthy people and people with diseases. Several articles have been published with data from the cohort.(23-25) The focus of the present study was symptom burden in people with multimorbidity and variables for the study was selected accordingly

Discussion
12. Page 11, lines 20 – 26 – again, please use “mental health diagnosis” instead of “mental diagnosis”
Thanks, the wording is now corrected.

13. Page 11, line 36 – rather than stating that patients become problem patients, please change this to be more patient-centered. Specifically, this could be stated as gradually becoming perceived by their clinicians as problematic patients, rather than those with problematic human situations.
Thank you for the suggestion; we have now removed the sentence.

14. Questionnaire – Why was information regarding number of sexual relationships gathered for women only and not men? This seems very problematic and also irrelevant to the research question as far as I can tell based on what is reported.

Please see our answer to your general comments above regarding the Danish Symptom Cohort. The purpose of the questionnaire was to collect information on symptoms and health care seeking behaviour in a broad perspective. We agree that it may seem odd to collect information on sexual relationships for women and not for men. However, among several aims with the cohort, one was to estimate of a number of gynecological alarm symptoms of cervical, endometrial and ovarian cancer, and to explore predictive values for gynecological cancer. The information of sexual relationships was developed related to this research question. As the overall strategy was to only include items relevant for the planned research questions, information of sexual relationships for men was not included, as no a priori research question was raised regarding this issue.

15. Overall – The discussion section is lacking mention of treatment burden, as symptom burden and treatment burden often go hand in hand (see, Cumulative Complexity Model Shippee 2012),

especially as patients with increasing symptoms have an increased need to seek care for relief. I think the discussion of these interplays could be strengthened.
Thank you for this very central comment. Both treatment and symptom burden is very central in multimorbidity, and patients with high symptom burden, but even more patients with many diagnoses, are likely prone to get a higher treatment burden. The questionnaire does not contain information on treatment burden, but it could be very interesting to study how treatment burden is affected by multimorbidity. In the manuscript we have added the yellow sentence in the discussion, and added Shippee as a reference:

This study shows that a mean of one new symptom is added for each morbidity, indicating that the symptom burden of patients with multimorbidity may be substantial. This may help clinicians to understand why patients with multimorbidity struggle to recognise which symptoms to focus on (36) and also help us understand why patients with multimorbidity are sometimes overwhelmed by their symptom burden.(5, 13) With the number of morbidities they have to live with, multimorbidity can quickly become a complex issue for patients,(37) not only because of the burden of symptoms, but also because of the burden of treatment,(38) the organisational challenges,(38) and the demands from everyday life.(33)
Reviewer: 2
Reviewer Name: Daniel Solomon
Institution and Country: Brigham and Women's Hospital
Harvard Medical School
Boston MA
USA
Please state any competing interests or state 'None declared': None

Comments to the Author
The question is interesting and the methods are well described. It appears that the results are valid.

It is not clear to me that the findings are important for research or clinical care. Having more morbidities is bad. It is interesting that symptoms are supra-additive, but its not clear what the implications are.
Thank you for the many good comments. We respond to them one by one here:

1. Should certain morbidities be targeted to reduce symptoms?

   Yes, we believe that the implications could be that for some combinations of multimorbidity it seems like symptoms are central and since symptoms is related to multimorbidity and it is known to reduce quality of life it is important to be able to know what combinations to be aware of or have certain focus on. But the most important clinical implication as we see it is that multimorbidity leads to higher symptom burden, not because it is supra-additive, but because at least one new symptom is experienced per extra diagnosis. Furthermore, for every symptom there can be a hospital contact, leading to a treatment, leading to another diagnosis, which in turn leads to yet another treatment, maybe with unwanted side-effects, leading to another doctor contact and so on. Multimorbidity quickly gets complex. It is known that multimorbidity is related to high symptom burden, but not how.

2. Are symptoms a bad measure of disease implications?

   Symptoms are one of the most important implications for how to understand health and a driver for seeking health care. However, symptoms are common in the population. We believe symptoms is a rather good measure (using this definition of multimorbidity) since approximately one new symptom is experienced for every new diagnosis.

3. I would also wonder whether the 10 categories of morbidities are a good measure of multimorbidity. Many morbidities within the same category are very different and the morbidity categories do not take into account different severities.

   Thank you for pointing out the differences in the ten disease categories. We completely agree that some of the disease within the same group are different e.g. sensory disease. On the other hand research performed on multimorbidity is very different both in number and type of

conditions. We wanted to perform at the same time relatively simple definition of multimorbidity that at the same was able to grasp complexity. We included information from the Danish registers and there is very little information on severity to retrieve from the registers. That would be preferable. However, in studies on multimorbidity only 28% of the studies to some extent include information about severity and, if so, it is primarily when using either self-reports or medical records as the source of data. In Denmark there is no valid register for data in primary care why this information was not included. However, we believe symptoms is a rather good measure (using this definition of multimorbidity) since approximately one new symptom is experienced for every new diagnosis.

Reviewer: 3
Reviewer Name: Sarah Monsell
Institution and Country: University of Washington, USA
Please state any competing interests or state 'None declared': None declared

Comments to the Author
1. Abstract: The last sentence of the Results seems to convey the same information as the Conclusion.
You're right. As requested by the editor the abstract has been remodelled, and does no longer contain the redundant information.

2. Methods: How did those who responded to the survey differ (or not) from those who did not respond to the survey in terms of patient demographics? It seems that the authors should have the data to perform this analysis.

We have added the following sentence to the methods section:
The focus of the present study was symptom burden in people with multimorbidity and variables for the study was selected accordingly.(23) The median age of the participants in the Danish Symptom Cohort was 52 years (IQR 40-64) and for non-participants it was 50 years (IQR 36-67). The respondents were reasonably representative of the study sample, but non-responders were more often men, unmarried, with lower education, lower income level, and with a generally looser attachment to the labour market.

3. Methods: Please expand on the reason for excluding gender-specific symptoms. Was their prevalence relatively high or low? If substantial, separate models could be run for men and women given the large sample size, at least as a sensitivity analysis.

The main reason for excluding the sex-specific symptoms from the questionnaire was that they were treated differently in the questionnaire and may have gained different attention than the rest of the symptoms. In the analysis plan, before looking into the data, the authors agreed to exclude symptoms that fewer than 100 reported i.e. blood in vomit and haemoptysis. However, the prevalence of the sex-specific questions did not differ significantly from the rest, but since only about 50% can answer yes to particular symptoms prevalence would be artificially low. If included we would have had to perform analyses separately for men and women (with doubled number of figures and tables) and loss of statistical power (or we would have had to adjust the models) as a result. Furthermore, since the main focus of the article was the synergy effects we were free to exclude symptoms.

4. Methods: Am I correct in reading that interaction effects between two specific diagnosis groups were evaluated? How frequent was the presence of 3 or more diagnosis groups for an individual? I think you may need to add the caveat to your conclusion that this "beyond additive" effect applies only to those with two diagnosis groups, and that the effect of having a combination of 3 or more diagnoses was not evaluated.

Yes, it is correct that it is the interaction effect of two diagnosis groups we evaluated and you are right that we should specify that it holds for combinations of two only. We have added the following in the results section:
Participants with multimorbidity (confined to participants with two morbidities) were analysed for interaction effects regarding their number of symptoms, impairment score, and worry score.

5. Methods: I think Poisson regression would better fit the outcome of number of symptoms. Similarly, ordinal logistic regression is likely a better fit for the other two outcomes that have 4 possible levels.

Both Poisson regression and ordinal logistic regression have assumptions that often are untenable (Poisson regression assumes equal mean and variance, the most popular ordinal logistic regression model assumes proportional odds between the outcome categories). Furthermore, we rather use the same model, and consequently the same definition for synergy, for each of our three outcomes. Hence, we employ linear regression. While the outcome may not be normally distributed, the means of the outcomes in groups (of not too few patients) are approximately normally distributed (because of the Central Limit Theorem) and this is used in the inference.

6. Figures 2, 3, and 4: These are really nice figures. My only concern is that we can't see confidence intervals. Please spell out "Syn" and "Eff." It would also be helpful to describe the interpretation of the effect. For example, an effect size of 4 translates to an average increase of 4 symptoms for participants with those two diagnoses compared to those with no diagnoses. This would be especially helpful given the different scales in each figure (and potentially link function as referenced above).

Thank you for this very good point. We have added three figure legends for each of the figures were we have explained them. We have also added an example describing the interpretation of the effect. Please see answer to reviewer 1, question 8. Furthermore, we have added all confidence intervals for the three figures in three supplementary files. All effects with a p-value <0.05 are marked with a star in the figures.

7. Results: We need to see confidence intervals from the models.

We have added all confidence intervals for the three figures in six supplementary files. All effects with a significant p-value (<0.05) is marked with an asterisk in the figures.

8. Results: Did the authors consider running models were the number of diagnosis groups (regardless of which specific groups), adjusted for the same covariates, was regressed on the symptom outcomes? I think this type of model would be highly informative. It would allow us to see if the relationship between the number of diagnosis groups and outcome in groups larger than two. After seeing the results for two diagnosis groups, I am quite interested in knowing if this effect holds as the number of diagnosis groups increases.

Before this study we have worked a lot with the definition of multimorbidity and after a systematic review of the definition of multimorbidity and thorough work we landed on this particular definition including ten groups of diagnoses. However, after performing this study we have thought a lot on if we should break up the ten groups and have another approach, were we study single diseases. Probably we will further on, but it will require a thorough work of what diagnoses and how many to include, since we otherwise can risk to end up with small groups of multimorbidity combinations. But, we totally agree that it would be interesting to study groups larger than two. Furthermore, please see reviewer 1, question 10.

VERSION 2 – REVIEW

REVIEWER	Kasey Boehmer Mayo Clinic, Rochester, MN, USA
REVIEW RETURNED	17-Mar-2021

GENERAL COMMENTS	The authors have addressed all of the major concerns in my previous review. The only minor detail (not requiring re-review) that should be addressed prior to publication is in the new Table 1, it would be helpful if the authors added to the key: 1. Range of low/high assets in the key and note what currency the ranges are noted in.

	2. Range of incomes in the 4 income quartiles and in what currency. It also might be better to label them as first (lowest), second, third, fourth (highest) quartiles vs having two labeled middle.

REVIEWER	Sarah Monsell University of Washington
REVIEW RETURNED	02-Mar-2021

GENERAL COMMENTS	Thank you for addressing most of my concerns in this revision. I did not see a response to comments document so please forgive me if you have already responded to these remaining concerns: 1. Supplementary tables B, D, and F should be in the manuscript and not the supplement. This is the only place the authors can see the confidence intervals around the effect estimates, which is how they are drawing conclusions around statistical significance. It is important that these be included as primary documentation of findings. 2. Symptom burden should be analyzed using Poisson regression comparing rates instead of linear regression comparing means. You are counting the number of symptoms so this is a rate, not a measurement. The other two outcomes would also do better with ordinal logistic regression or even collapsing into binary logistic regression depending on the distribution of the data; this measure is not linear and is subject to floor and ceiling effects.