PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cumulative social disadvantage and hospitalisations due to ambulatory care sensitive conditions in Finland in 2011–2013: a register study
AUTHORS	Lumme, Sonja; Manderbacka, Kristiina; Arffman, Martti; Karvonen, Sakari; Keskimaki, Ilmo

VERSION 1 – REVIEW

REVIEWER	Simone Kiel Institute for Community Medicine, Department of General Practice, University Medicine Greifswald, Germany
REVIEW RETURNED	22-Mar-2020

GENERAL COMMENTS	This is a very intersting topic and based on individual data with a very large sample size due to registry data. Page 5, line 6: The sentence is missing the word "study". Line 7: Please do not start the sentence with: "And third..." Page 6, line 14, Question 1: What do you mean by 'independent effect'? Independent of what? You are measuring the association of social risk factors and hospitalization due to ACSCs. An association is never independent. Hospital destricts as an indicator of region of residence may not be the best indicator. It has some limitations. You are invited to discuss this in the limitation section. Page 9: Please define educational level more detailed. What is meant by comprehensive school? Is it 10 years of school? Page 10, line 8: I assume you checked for multicollinearity between the explanatory variables. But what is meant by 'modest correlations'? Can you please insert the correlation coefficient you found between the explanatory variables? Page 12, line 4 and 5: Can you please mention the corresponding percentage of how many were hospitalized due to ACSCs? I assume 50 121/1 530 397 (xx%) and 133 341/927 152 (xx%). Also, there is no information about characteristics of the study population. How many people were categorized as living in poverty, with low level of eductaion, as unemployed and living alone? Page 12 line 9: "All the studied social and socioeconomic risk factors had an independent effect ...." There is no such thing as an independent effect. But an association between the outcome

	(dependent) and the independet variables. Please change or delete the word 'independent' throughout the entire manuscript. Page 15: How many people were categorized with prolonged cumulative disadvantage considering the number of years?

REVIEWER	Laura Rosella University of Toronto, Canada
REVIEW RETURNED	05-Apr-2020

GENERAL COMMENTS	This study examines the relationship between types of social disadvantage using a population based social registry. This work builds from previous studies and benefits from a large database with comprehensive individual-level information. There are some methodological questions that could be clarified to strengthen the manuscript. 1. A table of the population characteristics at baseline with and without ACSC would be extremely helpful in the manuscript before the modelling results. This helps understand the population cohort characteristics before modelling. 2. Time should be more clearly specified in this study. Is this ACSC in X time period? Among those without history of ACSC? The study design as a result is not clearly described. Could the authors more clearly describe timing if the exposure and outcome? This would help clarify the appropriateness of the statistical analysis as well. A figure in the supplement would be even better to make the design and selection of the analytic cohort very clear to the reader. 3. My main concern around the analysis is the choice of statistical model. The authors should justify the use of logistic regression, versus log-binomial or Modified Poisson model, either of which would give direct measures of relative risk compared to an odds ratio from the logistic model, which is subject to misinterpretation and overestimation of risk. (See McNutt LA et al. Am J Epidemiol; Zou G. Am J Epidemiol. 2004; Knol MJ et al. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. CMAJ. 2012 May 15;184(8):895-9.) Furthermore, if a longer time frame is used, a survival approach may also be appropriate and help account for time at risk, which the logistic model does not. 4. Two recent studies that also examine social disadvantage (along with other factors) and ACSC that may be worth reviewing in the Discussion as they support the findings and use linked population level data - they include: (a) De Prophetis at al. BMJ Open 2020 https://bmjopen.bmj.com/content/10/2/e032837.abstract and (b) Wallar et al., 2020 PLoS One. https://journals.plos.org/plosone/article/comments?id=10.1371/journal.pone.0229465) 5. Typically, 75 years of age are used as an upper bound on the ACSCs – with the idea that avoidable hospitalizations above this age are interpreted quite differently. Can the authors justify the inclusion of all hospitalizations in all ages? – and comment further how this affects their findings? This in fact is an alternate explanation to the 65+ findings.

	6. Although the authors did have individual level data, there were limited confounders available for control. This should be a more prominent point in the interpretation.

VERSION 1 – AUTHOR RESPONSE

Responses to the reviewer's comments

Reviewer 1
Page 5, line 6: The sentence is missing the word "study".

A: Thank you for noticing this, the word has now been inserted.

Line 7: Please do not start the sentence with: "And third..."

A: The sentence has now been edited.

Page 6, line 14, Question 1: What do you mean by 'independent effect'? Independent of what? You are measuring the association of social risk factors and hospitalization due to ACSCs. An association is never independent.

A: We agree that the formulation was ambiguous and we have now rephrased it:
What is the univariate effect of each risk factor on the risk of being hospitalised due to ACSCs if individual has no other risk factors?

Hospital destricts as an indicator of region of residence may not be the best indicator. It has some limitations. You are invited to discuss this in the limitation section.

A: We agree with the reviewer that the choice of indicator of region of residence is not easy. There are several options when defining the region of residence in Finland, depending what one aims to study. The main idea in taking into account the region in this study was to adjust for the differences seen in the incidence of ACSC hospitalisations between regions. We decided to use hospital districts since the distribution of variance in ACSC hospitalisations was was mainly distributed to the hospital district level and not to the health centre area level (Manderbacka et al., 2019). Moreover, the number of municipalities and health centre areas is high and using municipalities or hospital districts as a covariate variable in these analyses would have not given any additional value. The results were similar when using the health centre division or municipalities; the estimates differed only in the second or third decimal place. Owing to this comment, however, we feel that it is necessary to modify the methods section a bit. Concerning the region of residence, we have now added some description as follows:

We used 20 hospital districts, based on an administrative division of the Finnish hospital care system, as an indicator of region of residence to adjust for the differences in the incidence of hospitalisations due to ACSCs between regions.

(Manderbacka K, Arffman M, Satokangas M, et al. Regional variation of avoidable hospitalisations in a universal health care system: a register-based cohort study from Finland 1996−2013. BMJ Open 2019;9:e029592. doi: 10.1136/bmjopen-2019-029592).

Page 9: Please define educational level more detailed. What is meant by comprehensive school? Is it 10 years of school?

A: We have now given more precise definition of this indicator. Comprehensive school is 9 years in Finland.

Page 10, line 8: I assume you checked for multicollinearity between the explanatory variables. But what is meant by 'modest correlations'? Can you please insert the correlation coefficient you found between the explanatory variables?

A: Yes, we have calculated correlation coefficients (and other measures of associations). The Pearson correlation coefficient (and Cramer's V and Phi co-efficient which gave identical results) was 0.06-0.34 among younger men and 0.06-0.29 among younger women. The highest coefficients were between living alone and poverty (men) and unemployment and poverty (women). Among older, the values were 0.09-0.37 (men) and 0.08-0.45 (women). The association was strongest between poverty and living alone. We have now added these values into the text.

Page 12, line 4 and 5: Can you please mention the corresponding percentage of how many were hospitalized due to ACSCs? I assume 50 121/1 530 397 (xx%) and 133 341/927 152 (xx%).

A: This is important information, thank you for noticing this. We have added these proportions by gender and modified the text.

Also, there is no information about characteristics of the study population. How many people were categorized as living in poverty, with low level of eductaion, as unemployed and living alone?

A: We have now added new table showing percentages of people categorized according to risk factors and ACSC hospitalisations.

Page 12 line 9: "All the studied social and socioeconomic risk factors had an independent effect ...." There is no such thing as an independent effect. But an association between the outcome (dependent) and the independet variables. Please change or delete the word 'independent' throughout the entire manuscript.

A: Thank you for this note. By using this term independent we tried to describe the method we used. The independent effect refers only to effects of risk factors, not to the association of risk factors and the outcome. In these independent analyses, we have categorised the risk factor variables in a way that if a person had only that risk factor present and not the other risk factors, he/she was categorised as having it. This enabled us to study merely the effect of that factor on the outcome. However, we see now that the term is not clear enough and we have modified the text. We changed the word independent to univariate.

Page 15: How many people were categorized with prolonged cumulative disadvantage considering the number of years?

A: We provide now the number of people in these different categories in the table.

Reviewer 2.

1. A table of the population characteristics at baseline with and without ACSC would be extremely helpful in the manuscript before the modelling results. This helps understand the population cohort characteristics before modelling.

A: We agree with the reviewer and have now added a table describing the characteristics of the study population.

2. Time should be more clearly specified in this study. Is this ACSC in X time period? Among those without history of ACSC? The study design as a result is not clearly described. Could the authors more clearly describe timing if the exposure and outcome? This would help clarify the appropriateness of the statistical analysis as well. A figure in the supplement would be even better to make the design and selection of the analytic cohort very clear to the reader.

A: We thank the reviewer for this comment. The exposure period was 2006-2010 and the period of hospitalisations due to ACSCs was 2011-2013. In these main analyses, we did not take into consideration whether individual had a history of ACSCs. We performed additional analyses (sensitivity analyses) where we included only those individuals who had not a history of ACSC hospitalisations, i.e. we studied incident cases. We have now improved the methods section clarifying the study setting.

3. My main concern around the analysis is the choice of statistical model. The authors should justify the use of logistic regression, versus log-binomial or Modified Poisson model, either of which would give direct measures of relative risk compared to an odds ratio from the logistic model, which is subject to misinterpretation and overestimation of risk. (See McNutt LA et al. Am J Epidemiol; Zou G. Am J Epidemiol. 2004; Knol MJ et al. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. CMAJ. 2012 May 15;184(8):895-9.) Furthermore, if a longer time frame is used, a survival approach may also be appropriate and help account for time at risk, which the logistic model does not.

A: The reviewer is correct that in some cases, logistic regression is prone to overestimate the relative risk. This applies the cases where the incidence of outcome is common. In this study, the incidence of the outcome measure was rather low, especially among the middle-aged. Thus the difference between relative risk and odds ratio is rather small. We have now changed the analysis method to modified Poisson regression approach even though the estimates were slightly smaller than when using logistic regression.

The idea of using survival approach would be interesting if we have had a study design suitable for survival analyses, for example, time to event data including follow-up of patients from diagnosis until death/end of follow-up. In this study, our exposure measures cannot be modelled in that way since there are no specific time points for exposure events (the study subjects may have lived alone for one year, every now and then, 5 years or even for the whole adult life). Additionally, the complexity of the exposure variables (combinations of different risk factors) is another definite restraint for using survival analyses.

4. Two recent studies that also examine social disadvantage (along with other factors) and ACSC that may be worth reviewing in the Discussion as they support the findings and use linked population level data - they include:
(a) De Prophetis at al. BMJ Open 2020 https://bmjopen.bmj.com/content/10/2/e032837.abstract and
(b) Wallar et al., 2020 PLoS One.
https://journals.plos.org/plosone/article/comments?id=10.1371/journal.pone.0229465)

A: Thank you for noticing this shortage; we have now added these references in the manuscript.

5. Typically, 75 years of age are used as an upper bound on the ACSCs – with the idea that avoidable hospitalizations above this age are interpreted quite differently. Can the authors justify the inclusion of

all hospitalizations in all ages? – and comment further how this affects their findings? This in fact is an alternate explanation to the 65+ findings.

A: It is true that some studies have used this upper age limit (e.g. Wallar and Rosella, 2020). However, it seems that most of the earlier studies have not used this age restriction regarding ACSCs (Purdy et al., 2009; Billings et al., 1993; Agnus et al. 2019; Vuik et al., 2017; McCall et al., 2001; Paul et al., 2019, etc.). Especially, the UK definition which we have applied in this study does not use this age restriction. Thus, we feel the discussion of this issue is not necessary due to prevailing practices.

Purdy S, Griffin T, Salisbury C, Sharp D. Ambulatory care sensitive conditions: terminology and disease coding need to be more specific to aid policy makers and clinicians. Public Health. 2009;123(2):169-173.
Billings et al. Impact Of Socioeconomic Status On Hospital Use In New York City. Health Affairs 1993 12:1, 162-173
Billings J, Anderson GM, Newman LS. Recent Findings On Preventable Hospitalizations. Health Aff. 1996;15(3):239-249.
Agnus et al. Hospitalizations for ambulatory care sensitive conditions as an indicator of access to primary care and excess of bed supply, GMC Health Services Research, 2019.
Vuik SI, Fontana G, Mayer E, et al. Do hospitalisations for ambulatory care sensitive conditions reflect low access to primary care? An observational cohort study of primary care usage prior to hospitalization. BMJ Open 2017;7:e015704. doi: 10.1136/bmjopen-2016-015704
McCall N, Harlow J, Dayhoff D. Rates of Hospitalization for Ambulatory Care Sensitive Conditions in the Medicare+Choice Population. Health Care Financ Rev. 2001;22(3):127-145.
Marieke C Paul, Jan-Willem H Dik, Trynke Hoekstra, Christel E van Dijk, Admissions for ambulatory care sensitive conditions: a national observational study in the general and COPD population, European Journal of Public Health, Volume 29, Issue 2, April 2019, Pages 213–219, https://doi.org/10.1093/eurpub/cky182

6. Although the authors did have individual level data, there were limited confounders available for control. This should be a more prominent point in the interpretation.

A: We have now added more discussion concerning this limitation.

VERSION 2 – REVIEW

REVIEWER	Simone Kiel Institute for Community Medicine, Department General Practice, University Medicine Greifswald, Germany
REVIEW RETURNED	25-May-2020

GENERAL COMMENTS	Page 3, line 6 (Abstract): please delete the comma before the 'and'. Page 8, line 2 (Materials): Please revise the first sentence. It is not scientific language. e.g. ... was defined as the study population. Please include the headings: strengths and limitations as well as conclusions. "We" was used a lot in the Materials section. I'm not 100% sure if that's appropriate.

VERSION 2 – AUTHOR RESPONSE

Page 3, line 6 (Abstract): please delete the comma before the 'and'.

Answer: Thank you noting this.

Page 8, line 2 (Materials): Please revise the first sentence. It is not scientific language. e.g. ... was defined as the study population.

Answer: We have now modified the sentence.

Please include the headings: strengths and limitations as well as conclusions.

Answer: We have now included these headings.

"We" was used a lot in the Materials section. I'm not 100% sure if that's appropriate.

Answer: We have now modified the Materials section and use mainly passive voice.