Author’s response to reviews

Title: Potential gains in health expectancy by improving lifestyle: an application for European regions.

Authors:

Koen Füssenich (koen.fussenich@rivm.nl)

Wilma Nusselder (w.nusselder@erasmusmc.nl)

Stefan Lhachimi (stefan lhachimi@uni-bremen.de)

Hendrik Boshuizen (hendriek.boshuizen@rivm.nl)

Talitha Feenstra (t.l.feenstra@umcg.nl)

Version: 1 Date: 18 May 2018

Author’s response to reviews:

Dear editor, dear reviewers,

We would like to thanks the reviewers for their time and effort to read our manuscript, but especially for the valuable comments made. In response, we hereby submit a revised manuscript, taking the comments made by the reviewers and the Editor into account as much as possible. We have formulated a response for each of the comments listed below. The revisions to the manuscript are marked in red. We hope that with these changes, your concerns have been addressed adequately.

Kind regards,

On behalf of the authors,

Koen Füssenich

Reviewer #1:

This paper begins by noting the growing attention paid to healthy life expectancy as a summary measure of population health. It takes advantage of the natural laboratory represented by Europe.
The authors take their relative risks from the DYNAMO-HIA model. For those unfamiliar with this model, it might be useful to add a sentence about how these were derived.

- ‘We have added a line explaining this in our methods section, referring to the online available extensive reports regarding the estimation of relative risks for DYNAMO-HIA.’

This study uses a limited number of risk factors, but they are some of the most important. Perhaps the authors could comment on what others might feasibly be studied in this way, given the limitations of data and information on relative risks.

- ‘We agree that data as well as time available for the research project commissioned by the EU has limited the number of risk factors considered. Physical activity would be a very interesting additional candidate, however, its measurement is not so straightforward and, while available in our EB dataset, we were not sure it was sufficiently consistent to enable trustworthy analyses. Physical activity was not available in the Dynamo-HIA model. We added a few lines in the discussion section.’

An obvious limitation is the age of the data, with BMI from 2005. As these data were from Eurobarometer, they must be self-reported. This is something to discuss in the limitations (as they do for alcohol).

- ‘Self-reported data almost always has obvious limitations, and should be mentioned. However, this seems more of a problem for alcohol than for smoking and BMI. Surprisingly few of the survey participants fall in the ‘medium’ or ‘high’ drinking categories, while high BMI or smoking is reported much more frequent. We added a line on this to the discussion. Please note that since all our data sources use self-reported data, our Odds Ratios are consistent with our prevalences. We agree that 2005 is quite old and would have liked to have more recent data, but that was not possible when doing the research.’

My greatest concerns relate to their alcohol data and analyses. The measure of alcohol consumption is based purely on quantity, and does not take account of pattern. Again, given the marked differences in drinking patterns across Europe, this should be discussed as a limitation, especially as the pattern of drinking scene in the Baltic states and Poland will be associated with a high rate of sudden death due to injuries and arrhythmias while that observed in central Europe, for example in Hungary and Slovenia, will be associated with much higher rates of cirrhosis. This has obvious implications for a study of healthy life expectancy.

- ‘We agree with the reviewer that alcohol was the most difficult risk factor in our analyses. We have considered including binge drinking as a representant of drinking pattern in our analyses. However, considering drinking patterns in the DYNAMO-HIA framework would require relative risks for both mortality and disability. We added it as a limitation to the discussion.’
Linked to this, the failure to identify a higher rate of alcohol consumption in Eastern Europe in the Eurobarometer data is a cause for considerable concern, given that almost every other study finds this. This leads me to think that perhaps some more caution should be expressed about this data source, which includes only 1000 respondents in each countries, and 500 and the smaller ones, and is known to be subject to considerable sampling bias in some. Similarly, the strong U-shaped curve seen with alcohol might also raise some concern, as it is now well established that this is largely due to a selection effect, as alluded to in the discussion.

- ‘We fully agree that the levels of alcohol consumption found in the Eurobarometer contradict many other sources and are cause for concern, and have rewritten the discussion to express our concerns more clearly. This is also true for the odds ratio’s, which are also extensively discussed. However, we think it is better to present our findings and discuss their limitations, than to leave out the alcohol results altogether.’

While accepting the need for brevity, surely a little more detail is required on how they developed the odds ratios for health status. At least one of the data sources is longitudinal. Did they use that property?

- ‘The odds ratios were derived from cross-sectional analyses of SHARE wave 4 and the disability and the French Disability and Health survey Enquête ‘d’handicap et Santé. We did not conduct longitudinal analyses based on SHARE as we were interested in the effect of the risk factor on disability prevalence, as required by our application of DYNAMO-HIA, and to allow for a similar approach between the SHARE dataset and the Enquête ‘d’handicap et Santé. In addition, longitudinal analyses, selecting persons who are nondisabled at baseline would have reduced the number of subjects in the analyses substantially. More information on the odds ratio’s is now added to the appendix.’

I was expecting to see something more in the online material but it simply referred the reader to another report.

- ‘We thought it would more appropriate and faster to refer to the online report already available containing all extra information needed. However, we have now also added the relevant information on the estimation of the odds ratio’s between lifestyle and disability and the lifestyle prevalences to the appendix.’

Similarly, it was not at all clear how the relative risks for the three factors combined were estimated, as I would expect that this could be quite complex, especially, for example, with BMI and alcohol consumption where I can see lots of scope for confounding and non-linear interactions.

- ‘ORs and RRs for all three risk factors were estimated separately, however when doing so, the other risk factors were corrected for in the regression models. When applying these in the models, risks were multiplied assuming there were no complex interactions.'
While more complex interaction patterns might be worth investigating for the ORs, it is of little practical use for our current application: We would then also need information on joint prevalence in each country. We had no access to a data source with such information. Our current lifestyles prevalence estimates taken from different survey years of the EB, so we had no person level data to estimate joined prevalence.’

The reasons that certain countries were excluded from the analysis, shown in square brackets on page 9, will be obvious to European readers but not necessarily to others. A brief note might help.

- ‘Luxemburg was excluded since it is quite small. Croatia was excluded since we did not have access to sufficient data on this country, due to its very recent entrance to the EU. Norway and Iceland are not part of the EU. An explanation was added to the appendix.’

The observation that the gender gap and healthy life expectancy is much narrower than for life expectancy is actually well known and explains, in part, the common finding of worse self-perceived health among women in surveys, which is in part a function of survivorship bias.

- ‘As the reviewer mentions, the gender gap has been described extensively by others before. It should however be noted that this is not so much a finding of our research, but represents the input data directly. Our models focus on how much of the gap could potentially be reduced by a more healthy lifestyle, rather than describe the size of the gap itself.’

Overall, this study has quite a few important limitations, in relation to the data sources, the estimation of relative risks, and the prevalence of different outcomes. As such, the findings should be interpreted with quite considerable caution. Nonetheless, given the material available to the authors, they have done as much as is reasonably possible in my view. Thus, given the high policy relevance of the paper, with healthy life expectancy high on the political agenda, I believe that this paper makes a sufficiently important contribution, albeit with many caveats. However, these caveats really should be spelt out in a little bit more detail, as should some aspects of the methods. This would also suggest that they may wish to frame the paper as showing what could be done with much better data, rather than relying on the figures produced here.

- ‘We thank the reviewer for this balanced judgement, acknowledging the limitations and challenges of our project. We fully agree and have adjusted our conclusions accordingly. While the data used in this study has many limitations, we believe it is important to show what is, and what is not possible given the currently available data. More methodological articles describing what can be done with models such as DYNAMO-HIA in general are also available. For all these models, data availability will be an important limiting factor.’
Minor point. LEPGH should be spelled out in full in the main text and not just in the abstract.

- ‘LEPGH is spelled out in full in the introduction, line 52.’

Reviewer #2: The article addresses a research question, which is of interest to science, individuals as well as the media: What can you do (on average) to improve your survival chances? Or: How many (healthy) life years can be gained by lifestyle changes.

The three lifestyles addressed are smoking, alcohol consumption, and being overweight / obesity.

As interesting as the paper is in itself, I think it requires some revisions and clarifications:

The main issue I see is already acknowledged by the authors: The rather negligible effect of alcohol on mortality (see Table 2). This is in stark contrast to what has been discussed in many papers, most recently probably by Trias Llimós et al. in IJE. [1]

As far as I understood, these risk factors were obtained from the DYNAMO-HIA model/software. It seems that those risk factors were not very good for the impact of alcohol on mortality. But how can we be sure that the risk estimates are better for the impact of other risk factors on mortality?

- ‘The relative risks used from the DYNAMO-HIA website are well documented and based on other literature. The relative risks do show an increased risk, J-shaped curve, of mortality when consuming alcohol. The very low impact of alcohol on mortality seems to stem from the low reported alcohol intake, which in our study was based on the EB datasource. Only a very limited number of people reported ‘medium’ or ‘high’ alcohol use, and this may well reflect underreporting. We added a clearer explanation of this to the discussion.’

Many estimates rely on data from surveys (e.g., EU-SILC) as input. Thus, we may have some issues with sampling size for the used estimates. Could you please address this issue by showing how much your final estimates could vary due to the uncertainty of estimates derived from (potentially) small samples? Or, alternatively, say something about the sample sizes used for the risk estimates and point out that it would not affect the final estimates.

- ‘The sample size of the Eurobarometer is small for each country. We have therefore pooled data over countries in order to estimate a more reliable age and gender pattern.,
The OR’s are based on SHARE dataset and the Enquête ‘d’handicap et Santé, which have much larger sample sizes, as reflected in the reported confidence intervals.

The final estimates are the combination of the lifestyle prevalence estimates, disability odds ratio estimates and the parameter estimates and assumptions made in the DYNAMO-HIA model. While possible in principle to investigate the combined effect of all parameter uncertainties using probabilistic sensitivity analyses, it would require a very substantial amount of extra analyses.’

Throughout the text BMI, overweight and obesity are somewhat used interchangeably. For instance, in the introduction the authors write that "[...] BMI, smoking and alcohol are among the most important risk factors ...". I would argue that BMI is not a risk factor in itself but rather being overweight or obese.

- ‘We have indeed not been strictly consistent in the use of these terms. We now refer to the risk factor as BMI, of which overweight and obesity are two of the levels, increasing risk.’

I also think that the categorization of BMI needs to be improved. The lowest category seems to be "lower than 25". It is well known that mortality is highly elevated if BMI is below 18.5 (common threshold for being underweight). Obviously, there is a selection effect (or reverse causation): The reason for a relatively low BMI could be pre-existing health conditions, of course. But this would apply similarly to zero alcohol consumption.

- ‘We agree with the reviewer that an BMI below 18.5 is associated with high mortality. For all risk factors a balance had to be found between being precise and being too detailed for the type of data available to us. More categories (or even continuous use of the data) would be preferable from the point of view of descriptive accuracy. A fourth category of morbid underweight would have too few data points to allow for proper estimation of Odds Ratios and prevalences. The categorization into three groups as used by us in the current paper is very common. Hence we choose to be consistent and apply the same groups.’

Smaller issues:

- The authors write that "[O]dds ratios were adjusted for age, sex, ...". How did you adjust for age (categorical variable, linearly, ...)?
- ‘In our analyses of French Disability and Health survey Enquête ‘d’handicap et Santé, we included age as a continuous variable in the model. If statistically significant (p-value < 0.05) age square was added to the model.

We checked for interactions between the risk factos and age, relative to model with only age (either age, or age + age square). For models with age square we checked the interaction with
age and age square combined, using the Wald statistic. In case we found a significant interaction, we checked again whether other interactions were needed.

In our analyses of the SHARE data, age was included as a categorical variable in the categories <65 and 65+. Interaction with a dummy for these categories was tested and when significant included. More information on the estimation of the odds ratio’s is added to the appendix.’

- Population data and death counts were used from two data sources. Why?
- ‘We obtained data on GALI prevalence, population counts and deaths from the Eurohex website in order to use the same data source as in the official HLY calculation of Eurostat. Only for a few countries these data were not available from this website and for these countries we used the Human Mortality database. Data on population counts and death are virtually the same in different the data sources.’

The authors write that a "pooled model across all countries was used..." What kind of model do you refer to?

- ‘The Eurobarometer data contained relatively few observations per country, while we needed estimates for each category, age and gender. Hence, data were pooled across all countries by estimating in a single model the relationship between risk factor prevalence as a dependent variable and age, gender and country as an explanatory variables using country dummies. The model used was a multinomial model, since its dependent variable were the different categories for each risk factor. For age the relationship can be complex and is nog simply caught by a linear model or interaction terms. Hence, to predict smooth estimates, a multinomial ‘vector generalized additive model’ was applied, using the R package ‘VGAM’.’
- At the end of the discussion section there is a "d" missing in "...compared to current drinkers."
- ‘Added’
- I think that the need for better data gathering on the European level is a good idea. The beginning of the sentence "We conclude that more research is needed" is not really necessary as this could be (and is often) written at the end of every research article. This does not add anything.
- ‘We removed this sentence’
- In the list of abbreviations a "s" is missing in "Assesment" –‘Added’
- The authors mention it only briefly that "it was assumed that smoking, BMI and alcohol were independent." Maybe the authors could point out how this might be not correct and what kind of influence it may have on their estimates (no quantitative assessment, of course).
A meta-analysis by Noble et al. (2015) shows that risky behavior tends to cluster, while there might also be a healthy cluster. To quantify this clustering is very difficult. Our EB data did not allow us to do so, since no single EB survey contained information on all three of our risk factors. See our response to reviewer 1, end of page 2 above how we dealt with this. Briefly, at the individual level, the existence of a positive correlation between lifestyles, would mean the potential effects of only improving one lifestyle factor would be overstated, as individuals improving their lifestyle would still suffer from a shortened (healthy) life expectancy from other risky behavior. However, our results are at the aggregate level and at this level, as long as risks are indeed approximately multiplicative, the influence is small. When risks would in contrast be more than multiplicative (e.g. drinking and smoking is more detrimental to health than a multiplication of the risks indicates) than our estimates would underestimate the benefits of reducing prevalence of a single risk factor. But also overestimate the benefits of reducing prevalence of several risk factors. In contrast when risks would be less than multiplicative removing a single risk factor is less beneficial, since the other risk factor would still remain and our study would overestimate benefits. It requires a more detailed analysis then we were able to perform to estimate such interactions among risks. We have added a few lines on this in the discussion.

A suggestion (not necessary since no forecasts are made in the article):

As far as I know, and in contrast to the impact of smoking, the peak of the impact of the obesity epidemic on health and mortality has not yet been reached. Maybe the authors can speculate how this might affect their findings if they repeated it some time in the future. The authors might also consider to discuss this in the framework of cohort effects (such as smoking) in their period mortality perspective.

- ‘The current study takes indeed a cross sectional perspective and hence is not able to analyze the influence of dynamics in risk factor prevalence. When smoking is decreasing, while obesity might still increase, it will imply that if we repeat our analysis over 10 years, the effects of smoking would be smaller and those of BMI would be larger. Hence for policy advice it would be relevant to make the caveat that we might overestimate the effect of smoking and underestimate the potential for intervention on BMI. We accordingly added a line in the discussion.’

Reviewer #3: The paper "Potential gains in health expectancy by improving lifestyle: an application for European regions" provides a very nice quantitative demonstration of the fact that the major part of differentials in healthy life years results from lifestyles and could thus be reduced by human action. This is an important message for health researchers, policy officials and public health experts. I like in particular the exploitation of different data sources. The authors did great efforts and I would be happy to see the paper published. I have only some comments and suggestions that I summarize below by paper sections.

*Introduction*

It is true that the EU describes HLY as an indicator for disabilities. However, given that the measure is based on the GALI indicator ("Global Activity Limitation Indicator") I think it would be more precise to characterize HLY as measure for activity limitations. Similarly, it might be more appropriate to substitute the term "health loss" used on the EU website by health "deterioration" or "decline".

- ‘We thank the reviewer for pointing out the sometimes confusing terminology, which is however as also indicated by the reviewer established EU terminology. Disability is often used as an umbrella term for various measures of functioning, including activity limitations. We prefer to stick to the EU conventions and use the term disability as HLY is generally known as a disability-free life expectancy.

For the same reason of consistency with EU terminology and because we use a dichotomous outcome disabled vs. non-disabled, we prefer using the term health loss rather than deterioration or decline.’

*Methods*

The text of this section provides a good and intuitive idea about data, analytic concept and estimation scenarios, but more details would be appreciated. The combination of different data sources requires usually several adjustments and kinds of "smoothing". I am aware that these are in most cases complicated and difficult to describe briefly in such a short paper. It is therefore a necessary consequence that the descriptions are not detailed enough to get a sufficient idea of
what has been done, even with the additional information in appendix and supplement. The reference to a report with 156 pages (Ref #9) is according to me not the appropriate way to solve this issue.

- ‘As mentioned to reviewer 1, we thought it would more appropriate and faster to refer to the online report already available containing all extra information needed. However, as a result of these comments, we have now also added the relevant information on the estimation of the odds ratio’s between lifestyle and disability and the lifestyle prevalences to the appendix.’

It would be good if data and R codes could be made available upon publication to enable interested readers to understand and reproduce the analysis.

- ‘Two of our data sources used are owned by others and only available upon request. Therefore, unfortunately, it will not be possible to make these available. However the DYNAMO HIA model, as well as the EB dataset are freely available already. The SHARE data can be obtained upon request by other researchers as well. Finally, our R code and DYNAMO applications was sent to the commissioner of this research, CHAFEA, from which it may be requested by other researchers as well. We would be quite willing to act as intermediaries in doing so.’

*Results*

Results are well explained and presented in tables. This is naturally the most precise way to present results. However, tables are at the same time probably the most unspectacular way to present results. I would therefore like to motivate the authors to think about alternative graphical ways to illustrate their main findings.

- ‘We have replaced the three tables in the main text by a figure, capturing the bulk of information in the tables, and moved the three tables to the appendix for completeness.’

With regard to the interpretation of the results, I missed the note that the different quantitative HLY gains of reduced smoking among women and men are most likely a result of the smoking epidemic process. The period approach used in the paper depicts women and men in different stages of the smoking epidemic. It is therefore highly likely that the impact of smoking will rise among women as well in the next 10-20 years. Making these characteristics of the period approach clear in the description of the results would be helpful to avoid misinterpretations by readers who are less familiar with the used estimation techniques.
We agree with the reviewer this might be an inevitable result of our cross sectional approach. However, due to different timing, we do not think that smoking in women will ever reach the levels of exposure experienced by men in the past, so that not all of the observed differences between men and women can be attributed to differences in timing of the peak of the epidemic, the level of the peak also differs. We added some discussion on the cohort effects to the discussion, see also our response to reviewer 2 who had a similar comment on page 6 of this rebuttal.

*Discussion and Conclusion*

The authors have nicely interpreted and discussed the results. Nonetheless, I think it is important to give more emphasis to the insecurities resulting from the combination of data from different sources - Dynamohia, SHARE, Eurohex, HMD, Eurobarometer - with their very different characteristics and weaknesses. Although I highly appreciate the use of information from different sources, such an analysis (with its methodological complexities, see above) naturally limits the validity and statistical power of the results. The subjective self-reports of survey respondents increase this insecurity. (The highest alcohol abstain rates among Eastern Europeans as reported by the authors are a nice example.)

We appreciate this comment. We have now very extensively discussed the limitations of our data in the discussion section. We do not feel that use of different sources in itself is more problematic than use of a single source. Also within a single source different interviewers or data-retrieval points may cause inconsistencies. For instance, within SHARE differences between countries will exist, and they might be just as large as the differences between SHARE data and EB data from the same country. No single source would have all the information needed in this research. The choice ofdatasources in this project was guided by an expert panel, who commented on an initial document with these and several alternative sources. In this way we tried to get the best data available and accessible at the time of the project. By combining sources in a deliberate way and using careful analyses, we were careful to be as consistent as possible, but acknowledge the inherent problems in our data.

The use of period outcome measures and the assumptions underlying the Sullivan method add even further limitations when it comes to drawing conclusions for the real populations. I recommend making these limitations and insecurities much clearer in the discussion and conclusion sections.

We rewrote the discussion with more focus on these aspects.
As mentioned above, the messages obtained from this study are highly valuable and very interesting for a broad audience, but the authors should try to avoid over-estimations of over-interpretations of their results.

Finally, I would like to note that I perceived the paper across broad parts somehow difficult to read. I therefore suggest revising the language to make the text more fluent and easily readable.

- ‘We carefully reread the paper and rewrote text to shorter sentences to improve readability. If the reviewers deem this necessary we could also ask for English language editing.’