Subject: Manuscript resubmission AOAS1602-020

Dear Dr. Griffin,

We would like to thank you for considering our manuscript entitled “Efficient estimation of age-specific social contact rates between men and women” (reference AOAS1602-020) for publication in the Annals of Applied Statistics (AOAS). It is good to hear that you think our work has potential is useful to applied statisticians.

We would like to thank you, the associate editor and the anonymous reviewer for the very helpful comments. We have revised our manuscript according to these comments and we feel this has substantially improved our manuscript. We would like to resubmit this revised manuscript for publication AOAS. The comments are addressed below in detail.

Thank you again for your consideration.

Sincerely,

Dr. J. van de Kassteele

National Institute for Public Health and the Environment - RIVM

PO Box 1

3720 BA Bilthoven

The Netherlands

E: jan.van.de.kassteele@rivm.nl

T: +31-30-274-3690

**Associate editor (AE) review of "Efficient estimation of age-specific social contact rates between men and women"**

This paper discusses the interesting and useful problem of estimating contact rates.

*We thank the AE for the encouraging comments. They helped us to considerably improve the manuscript. The AE's comments and concerns are addressed below.*

**Main comments**

Model choice: One model set-up for estimating the contact rates is presented but it is not clear to me that this set-up is the most appropriate one for the data at hand, in particular:

Penalization: A second order RW is used, why not a first order RW? A 2nd order RW implies an expected continuation of rates of change in contact rates across age groups, but it’s not clear to me that this is more appropriate for this application than a 1st order RW, whereby extrapolations are based on an expected continuation of the levels of contact rates across age groups. So I would encourage the authors to try out a 1st order RW as well and do model comparison (see next main point).

*This is a good point; we have now added a justification for the choice of a second order RW. Our choice for a RW2 prior follows common practice with spline smoothing, where it is common to penalize the second derivative of the curve (e.g. Wood, 2006), and P-spline smoothing (Eilers and Marx, 1996), where it is common to put a discrete second order penalty on the spline coefficients. In a Bayesian context, this corresponds to putting a RW2 prior on the coefficients (Lang and Brezger, 2004). To test whether this common choice is also appropriate in the present setting, we have examined the effect of putting RW1, RW2 and RW3 priors on the contact rates. We have computed the WAIC, an information criterion, to compare the models. We found the RW2 prior to have the lowest WAIC, and we use this as a justification for our choice of prior.*

*We have added the following text to the manuscript:*

*Methodology section: “We examine the effect of applying RW1, RW2 and RW3 priors. The models are compared in terms of the Watanabe-Akaike or widely applicable information criterion (WAIC) (Watanabe, 2013).”*

*Results section: “Table 1 shows the WAIC and effective number of parameters for models with three different RW priors. The RW2 prior resulted in the lowest WAIC, and is therefore to be preferred. The RW1 prior implies the least regularization and highest effective number of parameters, the RW3 prior implies the most regularization, the smoothest surface and lowest effective number of parameters.”*

*Discussion section: “Based on the WAIC, the RW2 prior is the preferred prior. In frequentist statistics it is common use to penalize the second derivative of the curve (Wood, 2006). Particularly in P-spline smoothing it is a common use to put a discrete second order penalty on the spline coefficients (Eilers and Marx, 1996). In a Bayesian context, this is exactly the same as putting a RW2 prior on the coefficients (Lang and Brezger, 2004).”*

Neighborhood structure: there seems to be missing edges in Figure 1, which would imply the absence of smoothing for some neighboring age groups. For example, for the male-male matrix, I would expect similarity between nodes 2 and 6 but there is no edge currently. Perhaps that’s because the illustration is for the use of a 2nd order RW, in that case, why not add the sequence 2-6-7? Similarly, for a 2nd order RW, do you not want to add 7-10-11? When using a 1st order RW, I’d expect the edges between 2 and 6, and 13-14 to be added.

*Thanks for pointing this out, this was not explained properly in the manuscript. We have now improved the description of the neighborhood structure. The missing edge in Figure 1 between node 2 and 6 (or node 13 and 14) is caused by the way the D matrix is constructed for the RW2 prior. We do this by taking the second order differences in both the horizontal and vertical dimension, and construct the D matrix for both dimensions by applying kronecker products. This is a simple and intuitive way of constructing a smoother in two dimensions. Because the shape of the graph is triangular instead of rectangular, it is impossible to construct second order differences near the diagonal.*

*One alternative would be to use the biharmonic operator, a discrete analogue of thin plate splines (Rue and Held 2005). This version of the 2-dimensional RW2 prior is included in INLA as the 'rw2d' latent model. Modifying this prior to a triangular graph turns out to be very difficult, and simulations showed almost identical results for this version as compared to our simple approach.*

*We have added the following text to the manuscript:*

*Discussion section: “Our 2-dimensional RW2 prior is based on the work by Currie et al (2002), where second order differences in both the horizontal and vertical dimension are taken simultaneously. Our approach is a simple and intuitive way of constructing a smoother in two dimensions for triangular shaped grids. (...) A 2-dimensional RW2 prior could be constructed by applying the biharmonic operator, a discrete analogue of thin plate splines (Rue and Held, 2005). Adapting this prior to a triangular grid as required here is challenging.*

Smoothing parameter tau: is it appropriate to assume the same tau for all matrices? What if contact rates among males vary more rapidly between age groups, as compared to among females or males-females?

Parameter beta (eq. 2.5): I assume that the same beta was used for MM, MF, FM and FF but that may not be appropriate?

Model fit/validity/performance needs to be checked and discussed. The authors include only a check for sensitivity to choices of priors, which is not sufficient to show that the model assumptions are appropriate for the problem at hand. One option for model checking is to carry out out-of-sample validation exercises, e.g. by leaving out random subsets of the data (participants).

Validation exercises could also be based on the leaving out of specific subsets, e.g. older ages, or all women within a 5-year age group, to check how well the model extrapolates to groups with more limited data, if the data used also have groups with more limited data (it would be informative to add a summary table to the appendix with the number of participants by age/sex group). Such exercise may be particularly useful to motive the choice of the penalization because we would expect differences in performance between 1st and 2nd order RWs.

*We agree that the manuscript lacked clarity on these issues and have adapted the text accordingly.*

*We focus in this manuscript on the mean contact rate, and treat tau and beta as nuisance variables. After talking to the researchers who collect these social contact data, we believe that it is reasonable to assume that the rate of change between ages for the contact process between men, between men and women, and between women is the same for all sexes. If this assumption would be the focus of our research, we could adapt the model and allow for the amount of smoothing to differ between the sexes, but this would at the expense of increased computational complexity. One could imagine to also make a choice between for example a RW1 prior for men and a RW2 prior for women, or to introduce adaptive smoothing, where tau depends on age.*

*We have added the following text to the manuscript:*

*Discussion section: “We have described the total number of contacts by a Negative Binomial distribution that is parameterized with a mean and a dispersion parameter. The mean is allowed to vary by age and sex. In this study we focus on the expected contact rates and therefore we treat the dispersion parameter theta and precision or smoothing parameter tau as nuisance parameters. If these parameters would be of direct interest they can be varied by age and sex, although this would be computationally challenging. We believe that it is a reasonable assumption that the dispersion and rate of change between ages for the contact process between men, between men and women, and between women are the same for all sexes.”*

*About the intercept. We have now tested whether it would be better take one beta (intercept) for all four matrices MM, MF, FM and FF (the triangular parts). We compared the WAIC of this model to a model with one global beta. We did this for the RW1, RW2 and RW3 priors. The WAIC did not improve by including different betas, and we have decided not include it in the manuscript.*

*The validation exercise is a very good point. We have computed the probability integral transform (PIT), which can be used as a Bayesian 'leave-one-out' predictive measure of fit. It is the value that the predictive CDF attains at the observation. If the observation is drawn from the predictive distribution, the PIT has a standard uniform distribution (Dawid, 1984). We used the nonrandomized version of the PIT that is tailored to count data (Czado et al., 2009). The results are very good: we have almost uniform PIT histograms. The PIT histograms of the several different models are added to results section.*

*We have added the following text to the manuscript:*

*Methodology section: “We compute the probability integral transform (PIT) to check the validity of the models (Dawid, 1984). (...) The PIT can be used as a Bayesian 'leave-one-out' predictive measure of fit or calibration check. If the observation is drawn from the predictive distribution, the PIT has a standard uniform distribution. The PIT is usually being visualized in histograms. We use the nonrandomized version of the PIT that is suitable for count data (Czado et al., 2009).*

*Results section: “Figure 5 shows the PIT histograms for three models with different RW priors. In particular the histograms for the models with RW2 and RW3 prior are nearly uniform, indicating a good model fit, i.e. the observations are drawn from the predictive distribution.”*

**Other AE comments (in order of appearance in text)**

Section 2.1 is very clear, but could perhaps be made more concise/less repetitive. It may also help readers if there is some convention used regarding the knowns and unknowns, e.g. greek letter for unknowns and roman for knowns.

In section 2.2, it is not necessarily clear what is going on with the MM, MF etc. E.g. in Eq. 2.5, it is not clear to me whether the same beta was used for all four combis or whether there is one beta for all. For vector x, I assume that all x’s are stacked from all age groups, but that may be better to write out as well to avoid confusion.

P.8 perhaps better to introduce new notation when defining the smaller versions of D1 and D2.

*Good to hear that section 2.1 is clear. We have revised it make it more concise. We now inform the reader that the hyper parameters are in Greek (e.g. beta, tau and theta). We have replaced mu by E(Y). The notation we use is similar to the common notation in the literature on contact patterns and on Gaussian Markov Random Fields. To avoid confusion, we added the MF superscript to all variables used in section 2.2. This helps to clarify that parameters beta, theta and tau are global parameters, and not sex specific parameters. We also have made a clear distinction between x and vec(x).*

The sensitivity to prior choice is checked using methods from Roos et al but it’s not explained what’s done. Perhaps add this analysis to the appendix?

*We have added explanation, and included this text in the new section on model choice and validation.*

*We have added the following text to the manuscript:*

*Methodology section: “The sensitivity to the hyper priors is examined by comparing the local change in the posterior parameter distribution to the unmodified posterior, in case the prior distribution is modified in a standardized way. This can be done without rerunning the model (Roos et al, 2015). The sensitivity is summarized by a single number, here the worst-case sensitivity, expressed as a percentage. Percentages above 100% indicate super-sensitivity.”*

There is some imputation when age is reported as a range or when there is preferential reporting to multiples of 5. Should this procedure be repeated to include the uncertainty associates with this missing-ness problem into the overall uncertainty assessment?

*This is a good point. We have now have applied multiple imputations to age and included the additional uncertainty associated with this missingness in our results.*

*We have added the following text to the manuscript:*

*Results section: “Some participants reported the age of contacts as a range. We multiple imputed (10 times) these records by uniformly sampling an age from that range. From age 20 onwards, the reported age of contacts showed a preference at ages that were multiples of five. To prevent spurious results, we corrected these ages by uniformly redistributing the peak in an age range between two years younger and two years older. The additional uncertainty associated with the multiple imputations are included in all results.”*

Section 3.2: define what crude rates are in this context.

*We have added the following text to the manuscript:*

*Results section: “The crude age and sex-specific contact rates c are shown in Figure 2. With "crude" we mean contact rates that are directy calculated from the data without applying any regularization. They are obtained by equation (2.3), where E(Y) is replaced by y.”*

Figs 2 and 3 are nice but do make it difficult to compare model fit to data. Would it be possible to add additional plots to get a more direct comparison, ideally one that also shows the uncertainty in the estimates and/or data?

*This is a good idea. It is difficult to visualize uncertainties in two-dimensional figures. We therefore have added figures of cross sections at some particular ages (both participants and contacts) in order to visualize the uncertainties associated with our estimates.*

*We have added the following text and figure to the manuscript:*

*Results section: ”Figure 4 shows cross-sections of Figure 3 for a male participant at age 10 and a female participant at age 40. The figure illustrates the uncertainties associated with the estimates. Contact rates for a 10 year old male with other males and females of different ages increase rapidly with age, having a sharp maximum with 10 year old males and females (classmates) and reaching another peak with 45 year old males (fathers) and 40 year old females (mothers). Contact rates for a 40 year old female with other males and females of different ages show a more gradual pattern. Note the discontinuity in the gradient at contact age 10 (10 year old males) and contact age 40 (40 year old females) as a result of the tailor made prior distribution.”*

P.13 last line, forgot something between 20 and 30.

*LaTeX issue. Corrected.*

P.15 this data -> these data

*Corrected.*

P.15 “we tested the validity”: as the authors point out themselves later in the text, this is not a validity exercise.

*Corrected.*

**Anonymous reviewer's comments**

This is an interesting article which describes a method to infer the rates of contacts between men and women of different ages. The hierarchical Bayesian method they have applied models the contact matrix. The data model is assumed to follow a negative binomial. The mean number of contacts for this distribution is modeled using a GMRF for which the imposed structure of the precision matrix effectively smooths the contact rates across ages as well as preserves the symmetry of the contacts rates. The authors also discuss an interesting application where the risk of respiratory infections may be determined by age and sex. My main comments are below.

*We also would like to thank the anonymous reviewer for the comments and suggestions. It helped us to considerably improve the manuscript. The comments and concerns are addressed below.*

I am not convinced that all non-symmetric dependent triplets are currently reflected in the precision matrix. For instance, in Figure 1, for the Female-Male and Male-Female matrices, the total number of dependent triplets is 18 (i.e. 12 + 6). I would expect a total of 30 dependent triplets, or 2n(n − 2), for the FM and MF combined, where n is the number of age categories. This could impact the difference matrix, and ultimately, the precision matrix.

*We are not entirely sure that we understand the reviewer's comments. More specifically, why there should be 2n(n – 2) triplets. We can explain why there are 12 (FM) and 6 (MF) triplets. First, we have to recognize that the entire matrix (that is MM, FM, MF and FF together) should be symmetric in the main diagonal from bottom-left (M1, M1) to top-right (F5, F5) as indicated by the colors. It is easy to see that the MM and FF matrices are symmetric. The FM matrix and MF matrix are symmetric too. This means that nodes 16, 21, 25, 28 and 30 in the FM matrix should return in the MF matrix, or else the entire matrix will not be symmetric. This explains why the number of triplets in the MF matrix is 12 and the number of triplets in the FM matrix is 6: in the MF matrix we do not need the diagonal. This is explained in section 2.3 in the manuscript.*

The prior for beta is not mentioned nor discussed in the text.

*This is a good point. We put a non-informative normal prior on the beta parameter.*

*We have added the following text to the manuscript:*

*Methodology section: “At the third level, hyper priors are specified for intercept, the precision (smoothing) parameter of the GMRF and the dispersion parameter of the Negative Binomial distribution. We place a Normal prior with mean 0 and precision 0.001 on the parameter for the intercept beta, a Gamma prior distribution with shape parameter 1 and rate parameter 0.0001 on precision parameter tau, and a Normal prior with mean 0 and precision 0.001 on the logarithm of the dispersion parameter theta: (...)”*

Figure 2 and 3 provide nice visual representation of the data and the model fit. It would be interesting to overlay the contour lines of Figure 3 to the crude contact rates of Figure 2. This would give a better idea of how well the model fits those data.

*We understand that adding contour lines to the crude contact rates of Figure 2 could give a better idea of how well the model fits these data. However, we prefer to keep these two figures separate. Instead, we have added a formal model validation to our manuscript in terms of WAICs and probability integral transform (PIT) histograms. Please see our comments to the associate editor. As can be seen in Figure 5, the results are very good.*

The model does not retrieve all patterns observed in the contact rates. For example, the higher contact rates of older men contact with younger women is not as clearly depicted on Figure 3. Explanation why this happens as well as explanation regarding any other discrepancies would be a nice addition.

*The reviewer has a good point here. Sometimes we would like to see things that perhaps are not there, or at least not as clear as we would like them to be. The 'older men with younger women' contact is one of them. We have revised the text accordingly.*

The authors discussed several circumstances where coarsening is present in the data. Did the participants also report rounded number of contacts by categories? If so, was anything done to address this issue?

*There is indeed coarsening present in the data. Most important are the age of contacts reported as a range and the so-called digit preferencing, where the reported age of contacts showed a preference at ages that were multiples of five. The participants were explicitly asked to report any contact they had during one day. We have checked if the number of reported contacts showed any preferencing, e.g. multiple of five, but this was not the case.*

*Concerning the reporting of age as a range and the digit preferencing in age, we now have applied multiple imputation, as suggested by the associate editor. Additional uncertainties associated with the coarsening are now included in all results.*

*We have added the following text to the manuscript:*

*Results section: “Some participants reported the age of contacts as a range. We multiple imputed (10 times) these records by uniformly sampling an age from that range. From age 20 onwards, the reported age of contacts showed a preference at ages that were multiples of five. To prevent spurious results, we corrected these ages by uniformly redistributing the peak in an age range between two years younger and two years older. The additional uncertainty associated with the multiple imputations are included in all results.”*