Letter referring to Van der Gaag et al. "A multistate model to project elderly disability in case of limited data"

Ralph Brinks, German Diabetes Center, Institute for Biometry and Epidemiology, Duesseldorf, Germany

The objective of the work of Van der Gaag et al. is to describe an estimator for the age-specific incidence and mortality rates of an irreversible disease (or state) in case the age-specific prevalence and overall mortality rate of a population is given. The authors use variations of an illness-death model (Figures 1 and 2 in Van der Gaag et al. (2015)), which dates back at least to Fix & Neyman (1951). The resulting estimator for the incidence and mortality rates is based on a three step approach, which the authors describe as being motivated by the works of Podgor and Leske (1986) and of Barendregt et al. (2003).

The problem we see in the article by Van der Gaag et al is that the properties of the new estimation methods have not been examined analytically nor evaluated in a simulation study. A simple but very important question is: how accurate does the method estimate what it is supposed to estimate?

We would like to contribute to this question and show that the new estimator leads to results which could be (i) made more accurate and (ii) obtained with less effort by another estimation method. A minimalistic setting for examining the accuracy of the new estimator is the question, how accurate the new estimator is able to reproduce the incidence from input data where the incidence (and mortality) that produced this input is known?

In this letter we restrict ourselves to the model in Figure 1 of Van der Gaag et al., which is the classical illness-death model without risk factors others than age. The notion is analogously to Van der Gaag et al.

First, we generate a prevalence data set based on known incidence and mortality rates. We choose the incidence rate $\theta(x) = (x - 55)_+/1000$ where t_+ means the positive part of t: $t_+ = t$ if $t > 0$ and 0 else. The general mortality μ_{tot} is chosen to be of Strehler-Mildvan type $\mu_{tot}(x) = \exp(-10.5 + 0.1\,x)$. Furthermore, we assume that the relative mortality r is 2 for all ages: $r(x) = 2$. If we assume no calendar time trends in the rates and no migration as in Van der Gaag et al., the age-specific prevalence $Q_D(x)$ is the solution of the following ordinary differential equation (Brinks et al. 2013):

$$dQ_D/dx = \{1 - Q_D\}\,\{\theta - \mu_{tot}\,Q_D\,(r - 1)/[1 + Q_D\,(r - 1)]\}.$$

Together with the initial condition $Q_D(55) = 0$, we can numerically calculate the associated age-specific prevalence.
If we then apply the new estimator to this prevalence (using the known μ_{tot} and r), we find that the mean relative error of the estimated incidence is 5.7%. The maximum relative error is 50%. If instead the estimator described in Brinks et al. (2013) is used, the mean and maximum relative error is 0.05% and 0.34%. Moreover, the estimator of Brinks et al. is easier to calculate, because it does not need any optimisation algorithms. The figure shows the estimated age-specific incidence rates compared to the incidence used as input

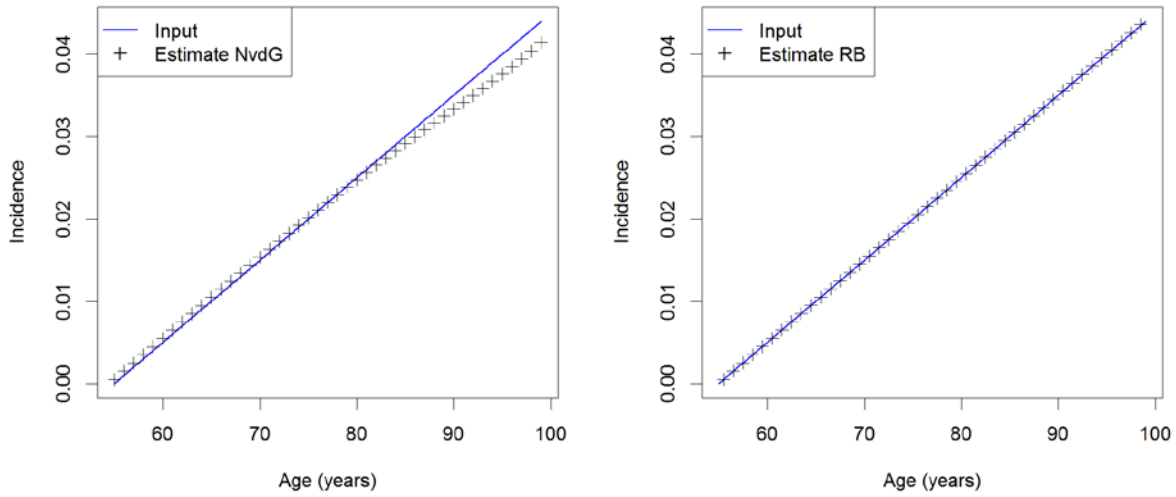

Figure: Estimated age-specific incidence (crosses) compared to the incidence used as input for the prevalence data (blue line). The left part of the figure shows the estimate by Van der Gaag et al. The right part represents the estimate of Brinks et al. (2013).

Hence, we would recommend to use the estimator by Brinks et al. which has also been generalised to cope with migration (Brinks & Landwehr 2014) and calendar time trends (Brinks & Landwehr 2015).

We appreciate the efforts of Van der Gaag et al. and thank the authors for their valuable work. Similar to Van der Gaag et al. we provide the R code for this comparison on the website of *Demographic Research*.

The source code can be downloaded here.

References

Barendregt, J. J., Van Oortmarssen, G. J., Vos, T., and Murray, C. J. (2003). A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. *Population health metrics*, *1*(1), 4.

Brinks, R., Landwehr, S., Icks, A., Koch, M., and Giani, G. (2013). Deriving age-specific incidence from prevalence with an ordinary differential equation. *Statistics in medicine*, *32*(12), 2070-2078.

Brinks, R., and Landwehr, S. (2014). Age-and time-dependent model of the prevalence of non-communicable diseases and application to dementia in Germany. *Theoretical population biology*, *92*, 62-68.

Brinks, R., and Landwehr, S. (2015). A new relation between prevalence and incidence of a chronic disease. *Mathematical Medicine and Biology*, doi: 10.1093/imammb/dqu024.

Fix, E., and Neyman, J. (1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology*, 205-241.

Podgor, M. J., and Leske, M. C. (1986). Estimating incidence from age‐specific prevalence for irreversible diseases with differential mortality. *Statistics in medicine*, *5*(6), 573-578.

Van der Gaag, N., Bijwaard, G., De Beer, J., and Bonneux, L. (2015). A multistate model to project elderly disability in case of limited data. *Demographic Research*, *32*(3), 75-106.