**Measurement error and correction of self-reported age in rural Malawi**

**Promotor:** prof. Wim Delva (Wim.Delva@ugent.be)

**Summary:**

The risk of HIV acquisition is believed to strongly depend on age as well as on the age gaps between an individual and his/her sexual partners. In an ongoing research, we are investigating the effects of sexual age mixing patterns on HIV acquisition risk in a well-documented rural community on Likoma Island in Malawi. However, our investigation relies on self-reported age, and strong evidence suggests that the ages among rural populations in sub-Saharan Africa are often estimated in an imprecise way, with great potential for bias [1]. This is due to the fact that these populations do not necessarily estimate age based on a precise data recording system, but rather employ a crude estimation approach, based on life events and physical well-being. As a result, digit preference around ages that end in “0” and “5” is common (known as age-heaping). Age-heaping can lead to attenuation bias, which means that the estimated effects of age and age gaps (and other co-variates that are correlated with age or age-gaps) on the health outcomes of interest, are biased when applying standard statistical regression methods.

Various methods have been developed to undo the attenuation effect of measurement error and thus reduce the resulting bias in generalized linear regression analysis [2,3]. Standard correction methods seek to estimate a correction factor (lambda), based on repeated measurement data. These repeated measurements are used to estimate the variability in the predictor variable (age in our case). The availability of a “gold standard” in the form of the birth register on Likoma Island offers a unique opportunity to precisely characterize the distribution of measurement errors (i.e. the differences between self-reported and true ages). Whereas standard methods may produce less biased regression estimates, they add uncertainty (wider confidence intervals) to the results, because of the uncertainty in the estimated correction factor lambda. In our proposed analysis, the measurement error will be known exactly for each study participant. This means we can investigate whether the measurement errors are truly random, as assumed by standard correction methods, or whether in fact, there are systematic trends that predict the size and direction of the errors. For instance, it is plausible that adults tend to overestimate their age if they report any illnesses that are perceived to be associated with elderly age. Conversely, childless women may tend to underreport their age, if social norms dictate that women above a certain age should be mothers.

Once the distribution of measurement errors is known, and demographic and health-related predictors of measurement errors are identified, regression models in which age and age gaps are dependent variables, can be modified to take into account this new knowledge. The self-reported age can be replaced by the true age, if the latter could be derived from the birth register. For individuals
that do not have a record in the birth register, an unbiased estimate of true age can be imputed, based on their demographic and health data. Comparison of regression analysis results from the naïve models (only self-reported age) versus the unbiased models (using true age and error-estimation sub-models) will provide a benchmark of the size of bias that may be introduced when estimating covariate effects in the presence of (non-)random measurement error in age among rural populations in sub-Saharan Africa.

References