ANCRTAdjust: An R package to adjust routine HIV testing data from antenatal care to reduce bias in estimating HIV prevalence trends

Brittany Blouin[1], Jeffrey W Eaton[2], Mary Mahy[3], and Mathieu Maheu-Giroux[1]

1 McGill University, Department of Epidemiology, Biostatistics, and Occupational Health **2** Imperial College London, Department of Infectious Disease Epidemiology **3** The Joint United Nations Programme on HIV/AIDS (UNAIDS), Department of Strategic Information

Summary

HIV surveillance in high-burden countries has traditionally been conducted using sentinel surveillance of pregnant women attending antenatal care (ANC-SS). Presently, administrative reporting data about routine HIV testing of all pregnant women attending antenatal care (ANC-RT) are increasingly being used for monitoring HIV epidemic trends (WHO, 2015). Importantly, using ANC-RT data could increase representativeness, geographical coverage, and granularity of information about HIV. If ANC-RT data on HIV serostatus are to be used for HIV surveillance, several challenges need to be addressed, however (Diaz, De Cock, Brown, Ghys, & Boerma, 2005). First, inferences from routinely collected program data could be biased due to imperfect data completeness, with some health facilities not reporting HIV testing data (WHO, 2013). Second, ANC-RT data could be inconsistent with missing and/or invalid data. Missing data problems include: missing entire reporting periods or missing one or more variables within a reporting period. Invalid data problems include: negative count values, the 'number of HIV-positive women' exceeding the 'number of women tested for HIV', and, the 'number of women tested for HIV' exceeding the 'number of women who attended the health facility'. Finally, imperfect testing coverage during antenatal care could result in selection bias, leading to overestimation of HIV prevalence (Maheu-Giroux M., 2019).

ANCRTAdjust is an R package that facilitates assessment of ANC-RT data quality, helps clean ANC-RT data, and, if warranted, can perform specific adjustments for common data inconsistencies as to minimize both information and selection bias in HIV prevalence estimates. Functionalities to longitudinally visualize and tabulate data quality indicators, as well as to identify outliers, are provided. A standard data cleaning function can be used to impute missing data (when possible), and to eliminate impossible negative count values and the possibility that the 'number of HIV-positive women' exceeds the 'number of women tested'. Specifically, different functions can be used to:

1. Evaluate data quality and completeness;
2. Perform data cleaning;
3. Identify outlying observations; and
4. Perform adjustments for common data inconsistencies due to incomplete reporting periods (using inverse probability of censoring weighting), multiple testing (occurring when the same women are tested multiple times during the reporting period leading to the 'number of women tested for HIV' exceeding the 'number of women who attended the health facility'), and/or imperfect testing coverage (using empirically derived adjustment factors).

As several countries transition from ANC-SS to ANC-RT data, it is imperative that clear and simple guidelines be established for the use of such data so that valid conclusions can be drawn for HIV surveillance purposes. The `ANCRTAdjust` R package was designed to be used by program managers, and public health practitioners and researchers alike, to help clean and adjust (if warranted) ANC-RT data for monitoring HIV trends in endemic countries. A standard protocol for data validation, implemented across all countries providing ANC-RT data for monitoring HIV trends, will ensure that HIV prevalence estimates are consistent, reproducible, and valid.

Acknowledgements

References

Diaz, T., De Cock, K., Brown, T., Ghys, P., & Boerma, J. (2005). New strategies for HIV surveillance in resource-constrained settings: An overview. *AIDS*, *19*, S1–S8. doi:10.1097/01.aids.0000172871.80723.3e

Maheu-Giroux M., K. T., Jahn A. (2019). HIV surveillance based on routine testing data from antenatal clinics in malawi (2011-2018): Measuring and adjusting for bias from imperfect testing coverage. *AIDS*, *In Press*.

WHO. (2013). *Guidelines for assessing the utility of data from prevention of mother-to-child transmission (PMTCT) programmes for HIV sentinel surveillance among pregnant women* (No. 9241505613). Wordl Health Organization, Geneva (Switzerland); Geneva, Switzerland.

WHO. (2015). *Conducting hiv surveillance based on routine programme data among pregnant women attending antenatal clinics*. World Health Organization, Geneva (Switzerland).