



Published in final edited form as:

Ann Epidemiol. 2022 January ; 65: 78–83. doi:10.1016/j.annepidem.2021.07.009.

Evaluating the quality of HIV epidemiologic evidence for populations in the absence of a reliable sampling frame: a modified quality assessment tool

Amrita Rao¹, Sheree Schwartz¹, Nikita Viswasam¹, Katherine Rucinski¹, Kimiko Van Wickle¹, Keith Sabin², Tisha Wheeler³, Jinkou Zhao⁴, Stefan Baral¹

¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St, Baltimore, MD 21205, USA.

²Strategic Information and Evaluation, Joint United Nations Programme on HIV/AIDS, Geneva, Switzerland.

³Office of HIV/AIDS, United States Agency for International Development, Washington, DC, USA.

⁴Technical Advice and Partnerships Department, The Global Fund to fight AIDS, Tuberculosis, and Malaria, Geneva, Switzerland.

Abstract

Background: Sampling frames rarely exist for key populations at highest risk for HIV, such as sex workers, men who have sex with men, people who use drugs, and transgender populations. Without reliable sampling frames, most data collection relies on non-probability sampling approaches including network-based methods (e.g. respondent driven sampling) and venue-based methods (e.g. time-location sampling). Quality of implementation and reporting of these studies is highly variable, making wide-ranging estimates often difficult to compare. Here, a modified quality assessment tool, Global.HIV Quality Assessment Tool for Data Generated through Non-Probability Sampling (GHQAT), was developed to evaluate the quality of HIV epidemiologic evidence generated using non-probability methods.

Methods: The GHQAT assesses three main domains: study design, study implementation, and indicator-specific criteria (prevalence, incidence, HIV continuum of care, and population size estimates). The study design domain focuses primarily on the specification of the target and study populations. The study implementation domain is concerned with sampling implementation. Each indicator-specific section contains items relevant to that specific indicator. A random subset of 50

Corresponding Author: Amrita Rao, arao24@jhu.edu, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St. Baltimore MD 21205, USA.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration: The views in this article are those of the authors and do not necessarily represent the views of USAID, PEPFAR, or the United States Government.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

studies from a larger systematic review on epidemiologic data related to HIV and key populations was generated and reviewed using the GHQAT by two independent reviewers. Inter-rater reliability was assessed by calculating intraclass correlation coefficients for the scores assigned to study design, study implementation and each of the indicator-specific criteria. Agreement was categorized as poor(0.00–0.50), fair(0.51–0.70), and good(0.71–1.00). The distribution of good, fair, and poor scores for each section was described.

Results: Overall, agreement between the two independent reviewers was good(ICC>0.7). Agreement was best for the section evaluating the HIV continuum of care(ICC=0.96). For HIV incidence, perfect agreement was observed, but this is likely due to the small number of studies reviewed that assessed incidence(n=3). Of the studies reviewed, 2% (n=1) received a score of “poor” for study design, while 50% (n=25) received a score of “poor” for study implementation.

Conclusions: Addressing HIV prevention and treatment needs of key populations is increasingly understood to be central to HIV responses across HIV epidemic settings, though data characterizing specific needs remains highly variable with the least amount of information in the most stigmatizing settings. Here, we present an efficient tool to guide HIV prevention and treatment programs as well as epidemiological data collection by reliably synthesizing the quality of available non-probability based epidemiologic information for key populations. This tool may help shed light on how researchers may improve not only the implementation of, but also the reporting on their studies.

Keywords

Quality assessment; Evidence synthesis; Non-probability sampling; Men who have sex with men; Female sex workers; People who inject drugs; Transgender populations; HIV

INTRODUCTION

In order to improve the reliability and utility of health research to inform policy and programs, a number of different reporting guidelines and quality assessment tools have been developed, including tools designed to evaluate randomized controlled trials and cohort studies^{1, 2}, measures of disease frequency (prevalence, incidence)³, reporting standards^{4–6}, and internal and external validity.^{5, 7, 8} Evidence-based decision making is predicated on a systematic assessment of both available information as well as the quality of that information, in order to determine which data points are best placed to inform next steps and which may be misleading. To date, there exists limited guidance on how to assess the quality of studies when non-probability sampling methods are employed.

Key populations, including female sex workers, men who have sex with men, transgender women, people who use drugs, and others, are an essential part of HIV surveillance and focus for HIV service delivery given proximal and structural risk factors that put them at greater risk for both HIV acquisition and onward transmission.^{9–12} Understanding the distinct HIV prevention and treatment needs of these groups is essential to develop responsive and effective programs and population-based interventions.^{9, 11, 12} Yet, the same factors that put key populations at greater risk for HIV acquisition and transmission and severely limit their access to prevention and treatment services – stigma, discrimination,

and criminalization – in turn make them harder to enumerate and characterize in a research context.¹³

While probability sampling approaches capitalize on random sample selection in order to protect against selection bias and ensure those selected for study are representative of a larger target population, there is often no way to construct a reliable sampling frame for key populations. Key populations may fear discrimination related to being identified as part of a stigmatized group and this may, in turn, prevent the enumeration of the total population of interest.^{13–15} This lack of a reliable sampling frame prevents researchers from using traditional probability sampling methods in studies of key populations. Many samples of key populations, then, are generated using non-probability sampling approaches including network- and venue-based methods.¹³ These methods, including snowball sampling, time-location sampling, and respondent-driven sampling, are well defined and understood.^{13, 16–18} The strategies are very different in their implementation and their potential for reaching a representative sample, and the appropriateness for the research question and the quality of implementation of these methods are rarely evaluated.

Existing guidance on statistical quality standards suggests that in order for a sample to be sound in quality, a probability sample is required. A report on non-probability sampling by the American Association for Public Opinion Research (AAPOR) noted¹⁹ that non-probability sampling methods have to undoubtedly overcome the challenge of selection bias (exclusion of large numbers of people, reliance on volunteers, and non-response), but not all non-probability sampling methods are equal when it comes to addressing the challenge of selection bias. For some methods, like convenience sampling, the only option may be to acknowledge that selection bias exists and add caveats to the interpretation of results accordingly. For other methods, like time-location sampling or respondent-driven sampling, the method itself makes it possible to generate weights and adjust the sample to approximate representation of the larger population.^{20–24} To further complicate the assessment, even when methods like time-location sampling or respondent-driven sampling could be used to better approximate the target population, poor implementation of these methods may render them no better than methods that do not involve weighting.^{20–24} Despite these differences across methods, little has been proposed about how to evaluate the relative quality of these studies.

In this paper, the Global.HIV Research Group presents a quality assessment tool that has been developed in order to consider the quality of epidemiologic evidence generated using non-probability methods. The National Heart, Lung, and Blood Institute's (NHLBI) Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies was used as a base for this tool.²⁵ The tool can be applied to the evaluation of the quality of key HIV epidemiologic indicators-- prevalence, incidence, the continuum of care and population size estimates – for key populations, in particular, and determines their relative utility for surveillance and decision-making.

METHODS

Development of the tool

The new tool, the Global.HIV Quality Assessment Tool for Data Generated through Non-Probability Sampling (GHQAT) was designed and draws some of its core components from The NHLBI Quality Assessment Tool of Observational Cohort and Cross-Sectional Studies.²⁵ First, the team consulted experts in systematic reviews on existing tools that may be used for the evaluation of key HIV epidemiologic indicators generated through non-probability sampling. When no suitable tool was found, a tool that flexibly evaluates the primary elements of observational studies was selected²⁵ and adapted to further assess specific methodologic concerns related to 1) non-probability sampling, 2) epidemiologic quantities used to inform decision-making, and 3) key populations. The adapted tool was piloted using real data to further refine it and account for challenges in practice and discrepancies between reviewers.

Structure of the tool

The GHQAT is made up of three main domains: study design, study implementation, and indicator-specific criteria. Each domain contains a set of items that together determine the quality score. The tool is designed to assess the quality of a given indicator, e.g. prevalence of HIV among female sex workers from a study in Malawi. The indicator was chosen as the unit of assessment rather than an overall study or a paper primarily because a study may report a number of different indicators (e.g. prevalence, treatment coverage, viral suppression, population size), but may only be designed to adequately capture one or some of those measures. It is possible, therefore, to have different quality scores assigned to different indicators produced by the same study.

Each of the main indicators (prevalence, incidence, continuum of care, population size estimates) receives a score that takes into account the quality of the 1) study design and 2) study implementation and reporting, which are both particularly important for studies that employ non-probability sampling. In each section of the tool, details and examples are provided regarding what is being asked or expected in each of the items to ensure consistency across reviewers.

Eligibility criteria and evaluation process

While the units of assessment used in this tool are the specific HIV indicators, multiple indicators may be captured within papers and there may be multiple papers published from any given study. Here, a study is defined as a particular sample of individuals among whom questions were asked or biological specimens were taken. Papers or reports are eligible for evaluation if they 1) had quantitative data of at least n=50; and 2) had data on any of the following indicators: HIV prevalence, incidence, continuum of care (HIV testing, diagnosis, treatment coverage, and viral suppression), or population size estimates. Papers or reports published from the same original study are evaluated together. If there are multiple papers published from the same study, a primary paper is selected for evaluation and the other published papers, often representing secondary analyses, are used to fill in any gaps in reporting. In doing this, greater priority is given to the quality of the data represented in the

indicator rather than the reporting in any given paper on that outcome. A paper is selected as the “primary paper” if it presented data that was the primary aim of the original study, if other papers published from the same study cite it for methodological details, and/or if it was published first. Both published academic papers and other data sources from the gray literature can be evaluated using this tool, including national surveillance system data reports, for example Demographic Health Surveys and Integrated Biological and Behavioral Surveys, as well as studies conducted by large international non-governmental organizations.

Inter-rater reliability and distribution of scores

To assess consistency across reviewers, inter-rater reliability was assessed for each item of the tool in a subset of 50 studies drawn from a global systematic review.²⁶ Eleven of the 50 studies included incorporate information from more than one paper (see Appendix Table 1 for details). Intraclass correlation coefficients (ICCs) for the scores assigned to study design, study implementation, and each of the indicator-specific criteria were calculated. All items had a yes/no response and were treated as dichotomous. Agreement was categorized as poor (0.00–0.50), fair (0.51–0.70), and good (0.71–1.00). Additionally, among the 50 studies evaluated for inter-rater reliability, the distribution of scores was plotted and described.

RESULTS

Developed Tool

The development of the tool involved an iterative piloting process prior to use. The tool, to reiterate, is made up of three domains. The indicator-specific criteria domain is further split into four sub-sections: prevalence, incidence, the continuum of care, and population size estimates, each of which are selectively completed based on whether or not a study has data on each of the indicators. The full tool can be found in the **Appendix**.

Study design—The study design domain is made up of four items and is designed to evaluate the quality of the study to answer the research question of interest:

1. Was the research question or objective in this study clearly stated?
2. Was the study population clearly defined?
3. Will the study population defined answer the research question proposed (adequately represent the target population)?
4. Was the sample size justified either through a power description, or variance and effect estimates?

Items one, two, and four of this domain come directly from the NHLBI tool. Item one assesses whether the goal of the study is clearly outlined, while item two asks whether or not a description of the study participants is provided, using demographics (person), location (place), and time period (time). Item four asks whether statistical power was considered in the selection of the study sample, or if selection was based primarily on convenience or cost. Item three was newly created to understand if those selected into the study (study population) would adequately represent the population for which inferences are to be made (target population). For example, if the authors are looking to understand viral suppression

among all female sex workers in an area, but the study design indicates only recruiting from clinics (where individuals are likely already linked to care), this study population may not adequately represent the target population.

Study implementation—The study implementation domain is made up of three items and is designed to understand how the sampling process may impact data quality. The items of this domain are:

1. Was the proportion of people who agreed to participate reported?
2. Was the proportion who agreed to participate at least 85%?
3. Is there reason to believe that the participants enrolled are a representative sample of the source population (the population from which participants were recruited)?

Item one of this domain was newly created and was added because many of the papers reviewed did not include information on proportion participating, defined as the number that participated out of the total number of those approached. Capturing information on participation, particularly in studies that employ non-probability sampling methods, shows a consideration for selection bias that is missed in those that do not. Item two is adapted from the NHLBI tool, which asks if at least 50% participate. Given that this tool is designed to assess the quality of evidence from a non-probability sample, representative participation from the source population is of even greater importance. A non-participation rate of greater than 15% was assessed to be too high, based on expert consensus among Global.HIV stakeholders. Scoring of the first two items depends in part on the study design. For example, reviewers are instructed that if dealing with a respondent-driven sampling study, they should look for participation rates or coupon return rates, but also consider whether assessments of equilibrium and/or bottleneck effects were conducted. If yes, the reviewers can use their discretion to assign the first two points for the study implementation domain. The first two items allow for some reviewer flexibility to consider participation rates and the larger context of representativeness and selection bias. For all other study designs, if participation rates are not reported, no points are awarded for either item, demonstrating the importance of reporting on representation. Item three assesses whether participants were selected with equal probability of sampling from the source population. For example, if a representative set of venues is chosen from which to recruit participants (the source population), are participants recruited equally from these venues? One issue, for example, could arise if participants who are known to be living with HIV systematically decline to participate when approached.

Indicator-specific criteria—Each indicator sub-section has items that assess quality of measurement, appropriate statistical analysis based on the indicator, the sampling approach to assess that outcome, and reporting. The incidence of HIV sub-section has additional items that assess the length of the study to assess incident infections, retention, and methods to address loss-to-follow-up. The HIV continuum of care includes an assessment of HIV testing, diagnosis, treatment coverage, and viral suppression, as a group, and points are awarded for each additional element of the continuum of care. The continuum of care

sub-section may be adapted if the goal of the assessment is to evaluate the quality of a single outcome within the continuum, for example looking exclusively at HIV testing or ART initiation. In this case, a point would be awarded for inclusion of that outcome and the indicator-specific score for the domain would be out of six rather than nine. In the population size estimates sub-section, additional elements around uncertainty reporting and implementation of the size estimation activity are assessed.

Scoring

There are three possible scores for each domain: “good”, “fair”, and “poor”. The score assigned to a given indicator is a composite of the individual items within the three scores: a study-design score, a study-implementation score, and an indicator-specific score. The study received 1 point from each “yes” to a given item. The scoring of each individual domain and sub-section is detailed in the first portion of Table 1. The final scoring for each indicator, based on the indicator-specific score+study design+study implementation, is detailed in the second portion of Table 1.

Inter-rater reliability and distribution of scores

On average, completion of the tool for a given paper takes approximately 15 minutes (+/- 5 minutes), with increasing time for those studies that have multiple relevant outcomes. The results of the assessment for inter-rater reliability are presented in Table 2. Overall, agreement between the two independent reviewers was good (ICC>0.7). The lowest ICC was for the prevalence sub-section (ICC=0.77). Agreement was best for the sub-section evaluating the HIV continuum of care (ICC=0.96). For HIV incidence, perfect agreement was observed, but this is likely due to the small number of studies reviewed that assessed incidence (n=3).

Of the 50 studies evaluated for inter-rater reliability, 20% (n=10) received a score of “good”, 78% (n=39) a score of “fair”, and 2% (n=1) a score of “poor” for study design. By comparison, over a third of studies (34%, n=17) received a score of “good” for study implementation, but half (n=25) received a score of “poor” for the domain. Studies performed similarly in the HIV prevalence and population size estimates sub-sections, while a greater proportion received a score of “poor” for the HIV care continuum sub-section. Too few studies (n=3) evaluated HIV incidence to adequately describe the distribution of scores (Figure 1).

DISCUSSION

Despite the variation in quality of different methods, there is currently no tool to assess the quality of data produced by non-probability sampling.¹⁹ The absence of a standard assessment of the quality of available data, limits the ability of policy-makers to weigh different pieces of evidence in order to make decisions. Indiscriminate use of data may misrepresent epidemic dynamics and can lead to assumptions about these key groups which may deflect or dilute resources away from where they are most needed. In order to address this gap, the Global.HIV Research Group adapted an existing quality assessment tool to evaluate the quality of HIV epidemiologic evidence, specifically for populations in which

no reliable sampling frame exists. In our assessment of 50 studies, we found this tool reliable and efficient to implement. Studies were most likely to lose quality points related to describing study implementation.

Unlike many of the existing tools, the GHQAT assesses data from both cross-sectional and longitudinal studies and can evaluate the quality of multiple HIV epidemiologic indicators. Additionally, it is easy to use, reliable, and transparent. Based on the ICCs generated for the different domains the reviewer agreement was shown to be good for the study design and study implementation domains. For the indicator-specific sub-sections, interrater reliability was shown to be close to perfect for the HIV care continuum, and good for prevalence and population size estimate sections. While it was noted that agreement was nearly perfect for the incidence sub-section, the small number of studies makes these results unreliable. It is recommended that, if possible, two reviewers evaluate the study in order to ensure consistent application of the tool. Given that the unit of assessment is the HIV-specific indicator itself, information about the study used to generate that data point (study design, study implementation, how different outcomes were assessed and reported) can be gleaned from different papers. Because of this, the tool is more forgiving of reporting omissions and focuses on awarding quality points when possible rather than solely penalizing based on lack of reporting.

This tool has three primary uses. First, it can be used by decision-makers, including program funders, implementing partners, and community organizations, for the thorough, rapid, and transparent assessment of the quality of available information. The tool may prove especially useful for decision-makers trying to demonstrate to constituents and others an evidence-based and impartial approach. The decision to use a particular data point or not will be dependent on the context and the availability of other data sources. The idea of the tool is to create a reference point for decision-makers to know if they can 1) use the data for planning with full comfort, 2) use the data with caution, or 3) advocate for and prioritize new data collection. Second, it can be used by those who are trying to synthesize information for the purposes of research, for example those conducting a systematic review, in order to comparably evaluate the quality of evidence using a single tool. Third, it can be used by those involved in the data generation and reporting process (i.e. researchers, program planners, NGOs) as a way to identify the key elements that should be included as part of both the implementation and reporting of a research study. Specifically, we see this tool not only facilitating the evaluation of the quality of existing studies, but also providing a guide for writing a good quality paper or report.

There are some important limitations of the GHQAT. While this tool allows for an assessment of the quality of evidence produced from studies utilizing non-probability sampling approaches, the content focus of this tool is really for HIV and key populations. For different research questions, one could arguably create new indicator-specific criteria sub-sections that are specific to another content-area and adapt this tool accordingly. Given that there are limited standard reporting guidelines for studies that employ non-probability sampling and the variety of types of studies we are hoping to evaluate (e.g. respondent-driven sampling, time-location sampling snowball sampling, etc.), we felt that some flexibility in evaluation was needed. We built in some flexibility in the items that

essentially requires a reviewer to use their discretion for certain sub-sections. Without this flexibility, we would have had to create a much longer, more cumbersome tool with sub-sections based on the method of sampling. To add to this, because papers from the same study can be evaluated together, use of this tool requires the reviewer to be able to determine if different papers are in fact from the same study. Sometimes this is a straightforward process, for example if the study has a recognizable name, but often, determining whether papers are from the same study requires the reviewer to triangulate different pieces of information, including sample size, geographic area, year(s) in which data were collected, and key authors. Given these complexities, use of the tool does require some training, especially in the organization of different papers from the same study and in the selection of a “primary paper.” These trainings, when conducted in the past, have been straightforward, with reviewers with some content-knowledge of HIV and key populations successfully completing quality assessment reviews using this tool. Based on our findings here and use of the tool, the items of the tool may continue to evolve. For example, it is possible that the 85% participation mark is too high and another round of expert input and consensus may be warranted if this is too infrequently met or unrealistic. Using the reviews performed so far as a guide, the larger issue observed is that studies provide few to no implementation details.

To date, this quality assessment tool has been used to evaluate over 250 studies, and has been presented to over 100 program implementing partners and policy-makers in four countries. This quality assessment tool has shown to be reliable and relatively easy to implement. Use of this tool can support in the evaluation of existing information, identify opportunities where new epidemiologic data may be needed, and provide policy-makers with a transparent approach to weighing different pieces of evidence in order to make decisions.

Acknowledgments

Funding: Amrita Rao is supported in part by the National Institute of Mental Health [F31MH124458].

Global.HIV Research Group

The Global.HIV Research Group is made up of a group of collaborators from the Johns Hopkins School of Public Health, The United States Agency for International Development (USAID), The Global Fund to Fight AIDS, Tuberculosis, and Malaria, and UNAIDS. The primary purpose of this collaboration has been to synthesize and make available and accessible all epidemiologic data related to key populations, in order to guide large-scale HIV funding and programmatic responses. Development of a quality assessment tool for these data arose from a need identified by key stakeholders and policy makers to better use and interpret available information. The methods for this global systematic review have been described previously.²⁶

References

1. The Joanna Briggs Institute. Reviewer’s Manual. The Joanna Briggs Institute; 2014.
2. Centre for Evidence-Based Management. Critical appraisal of a survey. <http://www.cebma.org/wp-content/uploads/Critical-Appraisal-Questions-for-a-Survey.pdf>
3. Loney PL, Chambers LW, Bennett KJ, Roberts JG, Stratford PW. Critical appraisal of the health research literature: prevalence or incidence of a health problem. *Chronic Dis Can* 1998;19(4):170–6. [PubMed: 10029513]
4. Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* Oct 16 2007;4(10):e297. doi:10.1371/journal.pmed.0040297 [PubMed: 17941715]

5. National Collaborating Centre for Environmental Health (NCCEH). A Primer for Evaluating the Quality of Studies on Environmental Health Critical Appraisal of Cross-Sectional Studies. http://www.nccceh.ca/sites/default/files/Critical_Appraisal_Cross-Sectional_Studies_Aug_2011.pdf
6. White RG, Hakim AJ, Salganik MJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology for respondent-driven sampling studies: “STROBE-RDS” statement. *J Clin Epidemiol* Dec 2015;68(12):1463–71. doi:10.1016/j.jclinepi.2015.04.002 [PubMed: 26112433]
7. Zeng X, Zhang Y, Kwong JS, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *J Evid Based Med* Feb 2015;8(1):2–10. doi:10.1111/jebm.12141 [PubMed: 25594108]
8. Hoy D, Brooks P, Woolf A, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. *J Clin Epidemiol* Sep 2012;65(9):934–9. doi:10.1016/j.jclinepi.2011.11.014 [PubMed: 22742910]
9. Baral S, Rao A, Sullivan P, et al. The disconnect between individual-level and population-level HIV prevention benefits of antiretroviral treatment. *Lancet HIV*. Sep 2019;6(9):e632–e638. doi:10.1016/S2352-3018(19)30226-7 [PubMed: 31331822]
10. Mukandavire C, Walker J, Schwartz S, et al. Estimating the contribution of key populations towards the spread of HIV in Dakar, Senegal. *J Int AIDS Soc* Jul 2018;21 Suppl 5:e25126. doi:10.1002/jia2.25126 [PubMed: 30033604]
11. Mishra S, Boily MC, Schwartz S, et al. Data and methods to characterize the role of sex work and to inform sex work programs in generalized HIV epidemics: evidence to challenge assumptions. *Ann Epidemiol* Aug 2016;26(8):557–569. doi:10.1016/j.annepidem.2016.06.004 [PubMed: 27421700]
12. Boily MC, Pickles M, Alary M, et al. What really is a concentrated HIV epidemic and what does it mean for West and Central Africa? Insights from mathematical modeling. *J Acquir Immune Defic Syndr* Mar 1 2015;68 Suppl 2:S74–82. doi:10.1097/QAI.0000000000000437 [PubMed: 25723994]
13. Rao A, Stahlman S, Hargreaves J, et al. Sampling Key Populations for HIV Surveillance: Results From Eight Cross-Sectional Studies Using Respondent-Driven Sampling and Venue-Based Snowball Sampling. *JMIR Public Health Surveill* Oct 20 2017;3(4):e72. doi:10.2196/publichealth.8116 [PubMed: 29054832]
14. Aday LA, Cornelius LJ. *Designing and conducting health surveys: A comprehensive guide* (3rd ed.). Josey-Bass; 2006.
15. Schwartz SR, Rao A, Rucinski KB, et al. HIV-Related Implementation Research for Key Populations: Designing for Individuals, Evaluating Across Populations, and Integrating Context. *J Acquir Immune Defic Syndr* Dec 2019;82 Suppl 3:S206–S216. doi:10.1097/QAI.00000000000002191 [PubMed: 31764256]
16. Hince ME, Miller WM, Bula A, et al. Achieving the first 90 for key populations in sub-Saharan Africa through venue-based outreach: challenges and opportunities for HIV prevention based on PLACE study findings from Malawi and Angola. *J Int AIDS Soc* Jul 2018;21 Suppl 5:e25132. doi:10.1002/jia2.25132 [PubMed: 30033589]
17. Zalla LC, Hince ME, Edwards JK, Michel J, Weir SS. The burden of HIV among female sex workers, men who have sex with men and transgender women in Haiti: results from the 2016 Priorities for Local AIDS Control Efforts (PLACE) study. *J Int AIDS Soc* Jul 2019;22(7):e25281. doi:10.1002/jia2.25281 [PubMed: 31287624]
18. Magnani R, Sabin K, Sidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS*. May 2005;19 Suppl 2:S67–72. doi:10.1097/01.aids.0000172879.20628.e1
19. Baker R, Brick JM, Bates NA, et al. *Non-Probability Sampling: report of the AAPOR task force on non-probability sampling*. 2013.
20. Barros AB, Martins MRO. Improving Underestimation of HIV Prevalence in Surveys Using Time-Location Sampling. *J Urban Health*. Jan 2 2020;doi:10.1007/s11524-019-00415-8
21. Karon JM, Wejnert C. Statistical methods for the analysis of time-location sampling data. *J Urban Health*. Jun 2012;89(3):565–86. doi:10.1007/s11524-012-9676-8 [PubMed: 22421885]

22. Sabin KM, Johnston LG. Epidemiological challenges to the assessment of HIV burdens among key populations: respondent-driven sampling, time-location sampling and demographic and health surveys. *Curr Opin HIV AIDS*. Mar 2014;9(2):101–6. doi:10.1097/COH.0000000000000046 [PubMed: 24464090]
23. Abdul-Quader AS, Heckathorn DD, Sabin K, Saidel T. Implementation and analysis of respondent driven sampling: lessons learned from the field. *J Urban Health*. Nov 2006;83(6 Suppl):i1–5. doi:10.1007/s11524-006-9108-8 [PubMed: 17058119]
24. Heckathorn DD. Snowball Versus Respondent-Driven Sampling. *Sociol Methodol* Aug 1 2011;41(1):355–366. doi:10.1111/j.1467-9531.2011.01244.x [PubMed: 22228916]
25. NHLBI. Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools>
26. Rao A, Schwartz S, Sabin K, et al. HIV-related data among key populations to inform evidence-based responses: protocol of a systematic review. *Syst Rev* Dec 3 2018;7(1):220. doi:10.1186/s13643-018-0894-3 [PubMed: 30509317]

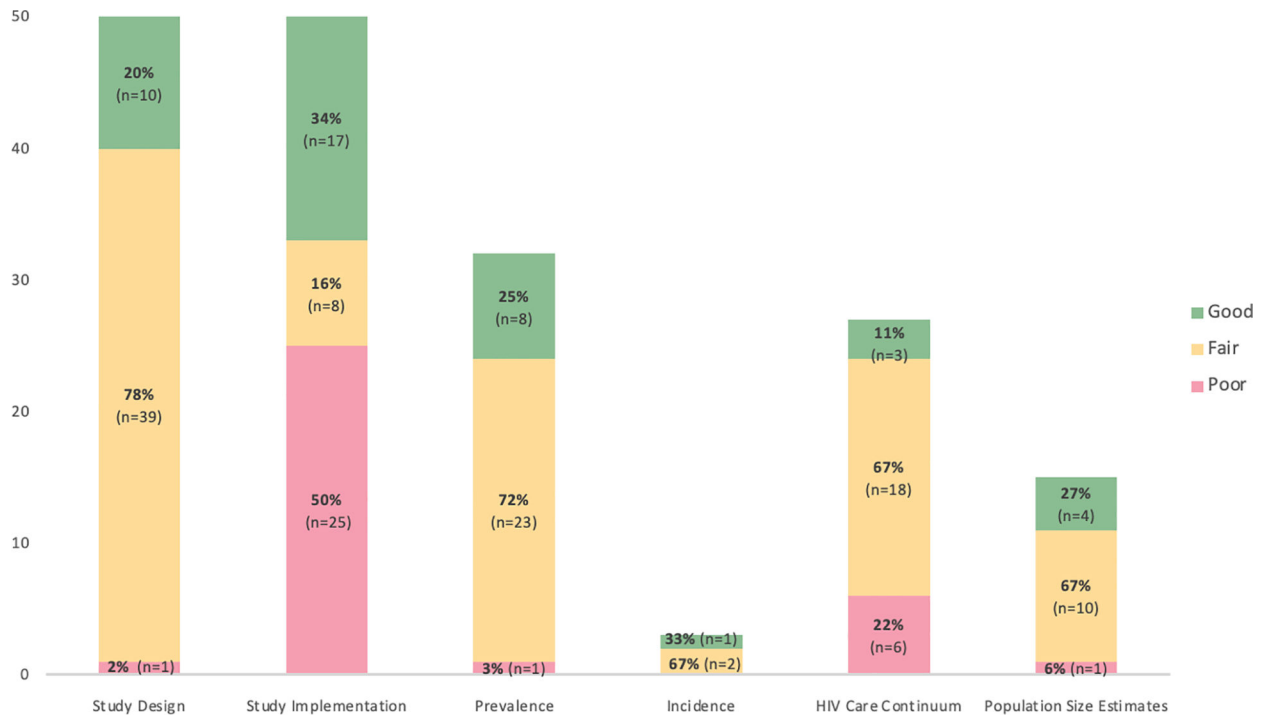


Figure 1. Distribution of scores for a random subset of studies (n=50) selected from the Global.HIV systematic review²⁶ using the Global.HIV Quality Assessment Tool for Data Generated Through Non-Probability Sampling (GHQAT)

Table 1.

Global.HIV Quality Assessment Tool for Data Generated through Non-Probability Sampling (GHQAT)
Scoring, By Domain and Overall

Domain/Sub-Section	TOTAL POSSIBLE PTS	GOOD	FAIR	POOR
Study design	4	4	2-3	0-1
Implementation	3	3	2	0-1
Prevalence of HIV	6	6	3-5	0-2
Incidence of HIV	9	8-9	5-7	0-4
HIV Care Continuum	9	8-9	5-7	0-4
Population Size Estimates	7	7	4-6	0-3
Domain/Sub-Section	TOTAL POSSIBLE PTS	GOOD	FAIR	POOR
Prevalence of HIV	13	13	7-12	0-6
Incidence of HIV	16	15-16	7-14	0-6
HIV Care Continuum*	16	15-16	7-14	0-6
Population Size Estimates	14	14	8-13	0-7

* includes estimates of HIV testing, HIV treatment, and viral suppression

Table 2.

Intraclass correlation coefficients (ICC) for inter-rater reliability of the Global.HIV Quality Assessment Tool for Data Generated Through Non-Probability Sampling (GHQAT) between two independent reviewers

	Study design (4 items)	Study implementation (3 items)	Prevalence (6 items)	Incidence (9 items)	HIV continuum of care (9 items)	Population size estimates (7 items)
Number of studies included	50	50	32	3	26	15
Intraclass correlation coefficient (ICC)	0.82	0.81	0.77	1.0	0.96	0.88
ICC 95% Confidence Interval	0.69–0.90	0.66–0.89	0.53–0.89	<i>Too few studies included</i>	0.90–0.98	0.64–0.96

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript