Interventions that aim to reduce risky sexual behaviour form the cornerstone of HIV/AIDS and sexually transmitted infections (STIs) prevention strategies in resource-constrained regions. Those working in prevention research need valid measures of HIV and STI reduction when developing and evaluating HIV/STI risk reduction programmes. ‘Biological measures’, such as incidence of HIV, are the most convincing indicators of programme impact. However, the large sample sizes required to demonstrate a reduction in incidence may exceed available research resources. ‘Observational measures’, such as observing used and un-used condoms in sex workers’ rooms, have been used to measure sexual risk reduction but can be challenging to implement.1 ‘Self-report measures’ are thus the mainstay of prevention evaluations and most HIV/STI prevention trials include self-reported sexual behaviours as a primary outcome. Research on the validity of self-reported measures is thus of fundamental importance.

The systematic review and meta-analysis by Phillips and colleagues, published in this issue of International Journal of Epidemiology, synthesizes the evidence from 15 studies comparing traditional face-to-face interview (FTFI) with innovative tools for recording self-reported HIV/STI risk behaviours.2 Innovative tools were developed to reduce the need for an interviewer, thereby increasing privacy and theoretically increasing reporting accuracy. This review focuses on studies from low- and middle-income countries, with most studies comparing FTFI with audio computer-assisted self-interview techniques. The authors found that, contrary to their expectations, FTFI was not uniformly inferior to non-interviewer techniques in reporting an increase in risky behaviours. They conclude that the direction of reporting is dependent on the sensitivity of the question and the population being assessed.

Use of systematic review methods to conduct a comprehensive search for eligible studies is essential for defining the current evidence base.3 In this review, the authors include both randomized controlled trials (11) and cross-sectional studies (4) presenting the crude estimates of effects combined for both types of studies. Tests for statistical heterogeneity due to study type were not significant but many would argue that it is inappropriate to combine the results from different study types despite the observed lack of significance.4 Instead, studies can be stratified by type and the estimates presented accordingly. Study quality is increasingly recognized as important in appraising the validity of study results included in a meta-analysis5 and an investigation of the underlying methodological quality of each included primary study would have greatly assisted the reader in interpreting these results.

In the absence of a gold standard reference test to measure sexual risk behaviour, no self-report tool can be proven to be more accurate than another as the real accuracy is not known. The primary studies included in this review are dependent on self-report behaviours and, as such, may be prone to bias. A review does not, of course, remove or reduce the potential biases inherent in the included primary studies, but it can make these explicit. Phillips and colleagues acknowledge this and state that their review is based on a number of assumptions. First, they assume that interviewees will generally under-report sexual behaviour to present themselves
in a more socially desirable light. It is just as reasonable to assume the opposite. Certain populations may regard sexual experience as prestigious and would arguably be more likely to over-report these behaviours.6,7 Secondly, their assumptions regarding the possible gender-driven directions of such reporting are bound by culturally determined norms. In direct contrast to accepted Western norms, a study in South Africa found that the dominant femininity incorporates tolerance of violence from partners.8

The complexities outlined above highlight the need for validation studies of self-report measures with biological tests as the reference standard. In a recent randomized controlled trial from Zimbabwe, 910 non-pregnant women were randomized to either FTFI or computer-assisted self-interview to elicit information regarding their recent sexual behaviour and condom use.9 All women then provided a vaginal swab which was tested for prostate-specific antigen (PSA) levels, a marker of recent semen exposure. Only 52% of PSA-positive women reported unprotected sex during the previous 2 days. In this context, self-report was a poor predictor of recent sexual activity with neither interview mode shown to be superior. Biological measures like these and others can serve as reference standards for recent unprotected sex but are only applicable for short, specific time periods of reporting and currently are only applicable to women.9,10 Despite these shortcomings, comparison studies between self-report measures and suitable references can help us understand the biases inherent in various forms of self-report measurement modalities, ultimately allowing better choices for outcome measurements.

Phillips and colleagues advocate for further research to understand factors affecting the degree of measurement error obtained with different interview methods given that different tools may generate different conclusions about the key components of HIV/STI prevention programmes. We would urge researchers and funders to carefully consider their chosen outcome measures. Randomized studies of comparisons between interview tools (FTFI or innovative) and valid reference tests of biological outcomes are urgently required. We may be surprised, as Phillips and colleagues were, to find that interviewers who convey a non-judgmental ethic and an ability to develop a rapport with respondents may be as accurate, or more so, than computers. Before such studies are approved, we would advocate that researchers conduct rigorous systematic reviews to identify and confirm the research gaps. To this end, Phillips and colleagues have made a good start.

Conflict of interest: None declared.

References