HIV Incidence Estimation Using Biomarkers for Recent Infection

Thomas Andrew McWalter

A thesis submitted to the Faculty of Science, University of the Witwatersrand, in fulfilment of the requirements for the degree of Doctor of Philosophy.

June 17, 2010

School of Computational and Applied Mathematics,
University of the Witwatersrand,
Johannesburg.

Declaration

I declare that this thesis is my own, unaided work, except where otherwise acknowledged. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other university.

_____

June 17, 2010

Abstract

Approximately one in six South African adults is infected with HIV, making it the country with the largest population of HIV positive individuals in the world. Strategies for monitoring this epidemic are an important area of research. In particular, estimation of incidence, the rate at which individuals are being infected, is a key indicator of the scale of the epidemic. Since it is cheaper, quicker, easier and potentially less biased than prospective follow-up, incidence estimation from the cross-sectional application of a biomarker that tests for recent infection has gained much attention. There is, however, controversy over how best to account for individuals that present anomalous biomarker responses. The central contribution of the thesis is to derive a consistent incidence estimation approach that accounts for anomalous responses. This approach is compared with other cross-sectional incidence estimators found in the literature and shown to be less biased. Implications of the new approach to survey design and the development of new biomarkers are explored. Application to survey data gathered by the Africa Center for Health and Population Studies showed consistent results when compared with incidence estimates derived from follow-up. Aside from other theoretical contributions, the thesis also provides a systematic review of the application of the BED assay in incidence estimation with recommendations on best current practice.

For my parents
David and Ingrid.

Acknowledgements

I would like to express my sincere appreciation to Alex Welte, my supervisor and friend. Our collaboration over the last few years has been enjoyable and vigourous, and, as a result, very productive. While this marks the end of your supervision, I am certain that we will continue to work together (hopefully in other domains of expertise too).

There have been a number of other active collaborators that I would like to thank: Till Bärnighausen, Tim Hallett, Reshma Kassanjee, Oliver Laeyendecker, Marie-Louise Newell and Zack Rosner. Your contributions are evident in these pages.

A number of other individuals who, while not being direct collaborators on the work presented here, have contributed in other ways to my interest and understanding in the field: Mike Busch, John Hargrove, Bill Rodriguez, Christine Rousseau, Brian Williams and everyone on the WHO Working Group on HIV Incidence Assays.

Finally, I would like to thank my family, friends and colleagues at Wits for support and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Prevalence and incidence are the two most important indicators of the state of an epidemic. Prevalence is the proportion of a population that has contracted an infection, while incidence is a measure of the risk of uninfected individuals contracting the disease and is usually expressed as a rate, i.e. the proportion of the at-risk (uninfected) population that become infected per unit time. Prevalence is the easier of the two indicators to measure, requiring only that the proportion of infected individuals be estimated by direct sampling in the population of interest. By contrast, considerable effort must be expended to estimate incidence.

In South Africa it is particularly important to have information on the state of the HIV epidemic since it has the largest population of HIV infected individuals of any country in the world (one in six South Africans aged 15-49 is infected [91]). Large-scale intervention is required to reduce the rate of new infections. The South African National Strategic Plan for HIV AIDS [95] has stated that one of its primary aims is to "Reduce the rate of new HIV infections by 50% by 2011." It is therefore necessary to have good estimates of incidence to ensure effective targeting and evaluation of interventions.

The most common way in which incidence is measured is by follow-up of an initially uninfected cohort. Over the duration of surveillance, individuals in the cohort are regularly tested, and incidence is estimated as the number of new infection events observed divided by the number of person-years of observation. The incidence estimated in this way is effectively an average incidence over the duration of the survey. Unfortunately, such longitudinal surveillance is expensive, logistically complex and prone to biases. These biases include the fact that certain individuals may become unavailable for follow-up, which may be correlated with risky behaviour, and that risk-reduction counselling, which must be extended to participants on ethical grounds, may affect behaviour during participation.

For infections with a relatively short duration (e.g. less than a year), another method for estimating incidence is available. By performing a cross-sectional survey of the population of interest, one can identify the number of individuals infected.

Incidence may then be calculated by inverting the well-known epidemiological relationship that prevalence is equal to incidence multiplied by the duration of infection. So, incidence can be calculated by dividing prevalence by the average duration of infection (given that an accurate estimate of the mean duration is available). Simplistically, this means that the difficult incidence measurement problem has been replaced by an easier problem of measuring infection prevalence. The incidence measured in this manner is effectively an average of the incidence over an historical period with length approximately equal to the duration of the infection.

Unfortunately, HIV has a long asymptomatic phase before the onset of immune failure and AIDS. This means that HIV infections last for many years and may not be diagnosed until long after the infection event. Furthermore, with the advent of antiretroviral therapy (ART), individuals who are enrolled on treatment programs may now survive for many decades. As a result, if one implemented the cross-sectional approach described above for HIV, the incidence estimate would be an average of incidence over decades of epidemic history. Such an incidence estimate would not be very useful. In the mid 1990s, however, a novel way of using cross-sectional surveys to estimate HIV incidence was proposed. The idea is to observe a biological marker (also known as a biomarker) indicating an immune system response to early infection and classify individuals as either recently infected or non-recently infected. Since individuals remain classified as recent infections by the biomarker for a much shorter period than they remain infected with HIV, the biomarker results can be used to estimate incidence in the same way that a short duration infection facilitates incidence estimation. An incidence estimate is computed in the same way as before, with one slight difference—the prevalence of recently infected individuals must be determined in the sub-population of the cross-section that excludes those that are non-recently infected.

Prior to conducting cross-sectional surveys to estimate incidence using a biomarker, it is necessary to estimate the mean duration that individuals spend in the recently-infected state. Since being classified as recent by the biomarker is sometimes referred to as "being in the window period", this mean duration is usually called the mean window period. Estimating (or calibrating) the mean window period requires longitudinal follow-up of individuals with approximately known infection dates. This requires considerable effort and cost, but need only be conducted once. Unless there is good reason to suspect that the mean window period is different in different contexts (e.g. as a result of sub-type diversity), all subsequent incidence surveys use the same value of the mean window period.

Brookmeyer and Quinn were the first to propose the biomarker-based approach in HIV monitoring [17]. They used the presence of p24 antigen (present in the

HIV protein shell) in a blood sample prior to HIV antibody production by the immune system (seroconversion) as an indication of recent infection. Unfortunately, the mean window period for this biomarker is very short (about three weeks), which means that to get good statistics, unreasonably large sample sizes for the incidence cross-section surveys are needed. Later, Janssen et al. [47] proposed a method based on the increase of a serological response (in particular they used 'detuned' assays to detect recently infected individuals). Depending on how the assays are applied, the mean window period for this approach is longer (between 100 and 200 days), facilitating better precision in the incidence estimates. This approach later became known as the Serological Testing Algorithm for Recent HIV Seroconversion (STARHS). Unfortunately, the use of detuned assays did not prove reliable due to the variability in immune response due to subtype diversity. In order to improve biomarker characteristics, a number of other assays that test for recent HIV infection have been developed, including the much used BED assay, which is a capture enzyme immunoassay (CEIA) based on protein sequences from the B, E and D HIV subtypes [79].

As the method of using biomarkers of recent infection in cross-sectional surveys was more widely applied, it became apparent that the results obtained from surveys invariably overestimated incidence. It was then realized that for the most useful assays (i.e., those with the mean window periods long enough for good statistics) a proportion of individuals remained perpetually classified as "recent" as a result of insufficient immunological response to HIV infection. The reasons for these anomalous responses are not completely understood, but it is well known that elite controllers (those individuals that have an innate ability to control the progress of HIV thus avoiding AIDS) fall into this category. It was also found that certain individuals on ART or with end stage disease experience immune system changes that may result in incorrect classification. These complications mean that the simple relationship between incidence and the prevalence of recently infected individuals is no longer obeyed. As a result alternative techniques for analysing data from such imperfect biomarker based surveys were needed.

To account for these anomalous results, McDougal and colleagues [63] at the Centers for Disease Control and Prevention (CDC), using statistical concepts from diagnostic testing, introduced an 'adjusted' incidence estimator which required not only the mean window period to be calibrated, but also a sensitivity and two specificity parameters (short- and long-term). This meant that, although imperfect tests like the BED assay could not be used for individual diagnosis, incidence could be estimated in surveys, provided that the calibration parameters were accurate for the population being surveyed. While this approach provided a way forward in analysing

data, it did so at the cost of requiring a complicated calibration exercise. To date, this calibration has only been carried out once in the context of North America.

Since this innovation, it was realized that the most important parameter for characterizing anomalous results is the long-term specificity parameter of the McDougal estimator. John Hargrove and colleagues [40] from SACEMA were the first to realize that the McDougal estimator is over-parameterised and provided a new estimator that depends only on the mean window period and a false-recent rate. The false-recent rate can be expressed as one minus the long-term specificity, and is essentially the proportion of non-recently infected individuals, infected for more than a certain time, that are incorrectly classified as recently infected. Later, Alex Welte and I [71] showed that an incidence estimator (different from that of Hargrove et al.), which also depends only on the mean window period and the false-recent rate, could be directly derived using a survival analysis of the problem. This estimator has been shown to be the least biased of the estimators available [69]. We were also able to derive a theoretically consistent way of reducing the number of calibration parameters in the McDougal estimator, in effect conclusively showing why their approach is over-parameterised [69].

An important consequence of the reduction of the number of calibration parameters in the McDougal estimator is that the complexity of the techniques required to calibrate the parameters that remain is less than the complexity required for the parameters that are eliminated. Another important fact that emerges from our analysis is that, under realistic conditions, the false-recent rate will vary with location and time. Our new estimator has been validated in an incidence study conducted in rural KwaZulu-Natal by the Africa Centre for Health & Population Studies [10]. This study showed that the local estimate for the false-recent rate was significantly lower than the equivalent false-recent recent rate found by McDougal et al. in the North American calibration exercise.

Recently, there has been some debate as to whether incidence estimates should be adjusted for false-recent results or not. Brookmeyer [15] has suggested that rather than using an estimator that accounts for false-recent results, one should consider all recent classifications as valid (i.e., within the window period) and use the simple estimator with a "better" estimate of the mean window period. While this approach is theoretically correct, it has some serious practical problems, including the fact that a calibration of such a mean window period must naturally happen over a very long period, since it must include a small proportion of individuals who may remain classified as recently infected for decades. Since such individuals are only removed from the population as a result of death, the calibration must be locally relevant. For example, it would be inappropriate to use a mean window period of

this type calibrated in North America for a study conducted in Africa, since access to basic healthcare, ART, nutrition and the mix of opportunistic infections in these two locations will lead to very different survival profiles for HIV infected individuals. Clearly, as these factors may change with time, the mean window period thus defined would also vary as a function of time.

While we disagree with the approach of Brookmeyer, the debate has, however, highlighted that there is a need for a more precise definition of the calibration experiments that are needed to characterise the performance of biomarkers. Further work in this direction is ongoing [65].

While there has been some controversy on how best to interpret data derived from cross-sectional surveys using biomarkers that test for recent HIV infection, the work undertaken has shown that there is a consistent, tractable and reliable method for producing incidence estimates using biomarker-based cross-sectional surveys. We continue to contribute to ongoing research in this area, and are also active in ensuring that these results are publicised to a wider audience of researchers and public health officials through various journal publications and contributions to the World Health Organization Working Group on HIV Incidence Assays. For example, we have provided a suite of spreadsheet tools for assay based incidence estimation (ABIE) available at `http://www.sacema.com/page/assay-based-incidence-estimation/`.

1.1 Contributions and Structure of the Ph.D.

In tackling this area of research, the primary goal has been to disseminate ideas as quickly as possible through peer-reviewed research articles. This has been achieved through the publication of five journal articles [71, 69, 118, 10, 8], two correspondence articles [117, 116] and a number of conference presentations (including [65, 70, 66, 54, 11]). These publications form a coherent body of work with a natural progression of ideas, and the chapters of the thesis reproduce the work roughly in the order that the material was conceived. The only modifications to the original papers have been some minor corrections, slight reformatting of the material and a unified set of references. At the beginning of each chapter a list of coauthors with whom the work was originally written and a reference to the original publication is provided. The structure of the rest of the Thesis now follows.

In Chapter 2 the original work on a consistent incidence estimator is presented. Here we derive a weighted incidence estimator that directly accounts for the phenomenon of assay non-progression (one source of anomalous results mentioned previously). We also explore the biases introduced due to simplifications made and derive an uncertainty relationship for the estimator that includes the effect of counting error

and parameter uncertainty.

Chapter 3 provides a rigorous comparison between the estimator derived in Chapter 2 and the estimators of McDougal et al. [63] and Hargrove et al. [40]. Initially, we prove a result that shows that, under their own assumptions, there is a redundancy inherent in the parameters required for the McDougal estimator. This result shows that the sensitivity and short-term specificity parameters are not needed for incidence estimation, and that a description of biomarker performance based only on the window period and long-term specificity (alternatively false-recent rate) provides an equally precise description of biomarker performance. The three paradigms are then compared under a model steady-state epidemic, with our estimator being shown to be the least biased of the three. This is consistent with the findings of Wang & Lagakos [112] who showed that this estimator is also the maximum likelihood estimator.

One of the consequences of the analysis in Chapter 3 is that the McDougal estimator can be simplified. In Chapter 4 we provide a short note that advertises this.

In chapter 5 we explore the consequences of the new framework for study sample size requirements, uncertainty and bias, under a variety of different scenarios. Given statistical requirements, we also provide an indication of the practical challenges that developers of tests for recent infection face. In an appendix to this chapter we show briefly how a general false-recent rate (as opposed to the probability of not progressing used in Chapter 2) can be used in conjunction with the incidence estimator. This means that, with a good estimate of the associated false-recent rate, it is possible to account for all the sources of anomalous results (i.e. assay non-progressors and regressors).

Work in this area would not be complete without at least one application of the statistical techniques to real data. In conjunction with our collaborators at the Africa Center for Health and Population Studies [2], we were able to show that the estimator performed well when compared to incidence derived from follow-up. One important finding in this work was that the false-recent rate for the population being studied was significantly different from other estimates of this parameter. This work is presented in Chapter 6.

There has been a progression in the sophistication of techniques used to analyse data from surveys that use biomarkers. Again, in conjunction with collaborators from the Africa Center, we performed a systematic literature review of the use and analysis of the BED assay in incidence estimation. Produced here as Chapter 7, there are a number of interesting findings that arose from the thirty nine studies reviewed. In particular, it was found that no less than ten distinct incidence estimators have

been used to analyse data from cross-sectional incidence surveys. Unfortunately, a large number of studies did not account for the presence of anomalous results, which may undermine the validity of their findings. Based on the findings in this work we provide recommendations on best current practice.

Recently Brookmeyer [15] has contested the need to "adjust" incidence estimates for anomalous results. In particular he suggests that the McDougal and Hargrove estimators do not increase the accuracy of incidence estimates, but that improved estimates of the mean window period are required. Chapter 8 reproduces correspondence in response to this work, in which we argue for the use of "adjusted" estimators.

In responding to Brookmeyer, a number of issues were raised—in particular, the need for a precise definition of the false-recent rate. Chapter 9 explores this in detail. By providing a precise characterisation of the calibration parameters (window period and false-recent rate) in terms of a predetermined cutoff time, a new estimator is derived. We provide an analysis of bias introduced as a result of simplifying assumptions, and compare the new estimator with previous estimators.

All of the graphs included in the Thesis were produced using MATLAB. In Appendix A, some examples of MATLAB code used are presented. Appendix B reproduces one of the conference posters presented at the International AIDS Society Conference held at Cape Town in 2009—it serves as an executive summary of Chapters 2-6.

Chapter 2

Relating Recent Infection Prevalence to Incidence with a Sub-population of Assay Non-progressors

∗ This chapter was coauthored with A. Welte [71], and is reproduced with permission from Springer Science+Business Media: Journal of Mathematical Biology (2010) 60:687-710 DOI: 10.1007/s00285-009-0282-7.

Abstract

We present a new analysis of relationships between disease incidence and the prevalence of an experimentally defined state of 'recent infection'. This leads to a clean separation between *biological parameters* (properties of disease progression as reflected in a test for recent infection), which need to be calibrated, and *epidemiological state variables*, which are estimated in a cross-sectional survey. The framework takes into account the possibility that details of the assay and host/pathogen chemistry leave a (knowable) fraction of the population in the recent category for all times. This systematically addresses an issue which is the source of some controversy about the appropriate use of the BED assay for defining recent HIV infection. The analysis is, however, applicable to any assay that forms the basis of a test for recent infection. Analysis of relative contributions of error arising variously from statistical considerations and simplifications of general expressions indicate that statistical error dominates heavily over methodological bias for realistic epidemiological and biological scenarios.

2.1 Introduction

Reliable estimation of disease *incidence* (rate of occurrence of new infections) and *prevalence* (the fraction of a population in an infected state) are central to the determination of epidemiological trends, especially for the allocation of resources and evaluation of interventions. Prevalence estimation is relatively straightforward, for example by cross-sectional survey. Incidence estimates are notoriously problematic,

though potentially of crucial importance. An approximate measure of incidence in a population is required for the proper planning of sample sizes and costing for clinical trials and other population based studies. Repeated follow-up of a representative cohort is often referred to as the 'gold standard' for estimating incidence, but is costly, time intensive and still prone to some intrinsic problems. For example, there may be bias in the factors determining which subjects are lost to follow-up. Furthermore, ethical considerations demand that a cohort study involve substantial support for subjects to avoid becoming infected, which may make the cohort unrepresentative of the population of interest.

Numerous methods [17, 26, 40, 47, 63, 79, 80, 83, 122] have been proposed for inferring incidence from single or multiple cross-sectional surveys rather than following up a cohort. A central idea in most of these [17, 26, 40, 47, 63, 79, 80] is to count the prevalence of a state of 'recent infection', which naturally depends on the recent incidence. The relationship between the two is in general not simple and depends in detail on the recent population dynamics as well as distributions which capture the inter-subject variability of progression through stages of infection, as they are observed by the specific laboratory assays used in the test for recent infection (TRI). For this approach to be sensible, a working definition of 'recent infection' must be calibrated, for example by repeatedly following up subjects over a period during which they become infected. This is effectively as much effort as one measurement of incidence by follow-up. The calibration can then be used to infer incidence from subsequent independent cross-sectional surveys.

Owing to the devastating impact of the HIV epidemic, and the many challenges of research and intervention design, the problem of estimating HIV incidence has attracted considerable interest in recent years. The prospect of using a TRI is in principle very attractive. Given the range of values of incidence likely to be observed in populations with a major epidemic (say 1-10% per annum) a mean definition of 'recent' of approximately half a year is desirable to yield reasonable statistical confidence for sample sizes of a few thousand. The BED assay[1] is currently the leading candidate for such a test, but controversy has arisen about the possibility of conducting a reliable calibration. This stems from the fact that a subset of individuals (approximately 5% [63, 40], potentially variable between viral and host populations) fail to progress above any statistically useful threshold set on the assay in the definition of 'recent' infection. This subset of individuals who remain persistently classified as 'recently infected', the so called *assay non-progressors*, poses a problem to which there is currently no consensus remedy. We emphasize that the analysis

[1] The BED assay is a capture enzyme immunoassay which uses protein sequences from HIV-1 subtypes B, E, and D [79].

performed in this paper is not only applicable to the BED assay, but applies to any assay that forms the basis of a TRI.

Of the references cited above, only McDougal et al. [63] and Hargrove et al. [40] have proposed approaches for addressing the issue of assay non-progression. In the former work four calibration parameters are used, being the window period (ω), the sensitivity (σ), the short-term specificity (ρ_1) and the long-term specificity (ρ_2) of the TRI. Some of these parameters are difficult to calibrate, requiring frequent and long-term follow-up. In an attempt to reduce the number of parameters, Hargrove et al. [40] provide a simplified formula under the assumption that $\sigma \approx \rho_1$.

We present a new analysis of the interaction between epidemiological trends and a model of inter-subject variability of progression through an experimental category of 'recent infection'. Our model yields simple formulae for inference *even when a fraction of the population fails to progress out of the recent category.* The only physiological assumption required to deal with the assay non-progressors is that their survival after infection is the same as the progressors. This assumption is also implicit in previous work on using TRIs to estimate incidence, as shown in [117, 67]. Our analysis shows that only two calibration parameters are required for the specification of the model, namely the proportion of seroconverters who do not progress above the threshold and the mean time it takes for progressing individuals to cross the threshold set by the TRI.

A key conceptual point about our analysis, which distinguishes it from all others of which we are aware, is that we confront the fundamental limitation of what can be inferred from a cross-sectional survey. In particular, even perfect knowledge of an instantaneous population state does not uniquely determine the instantaneous incidence. At best, a weighted average of recent incidence can be inferred. Although the discrepancies between this weighted average and instantaneous incidence can be shown to be small compared to statistical errors (for our application), it can in principle be systematically incorporated into estimation of trends from multiple cross-sectional surveys.

The article is organized as follows. In Section 2.2 we develop a basic continuous time model defined by a time dependent incidence and susceptible population, a distribution of times after infection spent under the threshold on a TRI and a distribution of post-infection survival times. We note that there is in principle no specific relationship between the instantaneous incidence and the prevalence of individuals who are infected and under the threshold. At best, one obtains a relationship between the prevalence of under-threshold individuals and a convolution of the recent incidence with a specific weighting function which is implied by the use of a TRI. This relationship in principle includes all moments of the distribution of the waiting

times that individuals spend under the threshold. We show that, for realistic rates of variation in the susceptible population, only the mean of the waiting time distribution is needed, and a simple expression for a weighted average of the incidence is obtained. The basic model is extended to allow some fraction of individuals (specified by a new parameter) to be assigned infinite waiting times under the threshold of the TRI. This leads to only very minor modifications of the previous expression for weighted incidence, namely a systematic 'subtraction' of over-counted 'not recently infected' individuals which are included in the experimental category 'under threshold'.

Section 2.3 explores the consequences of designing a cross-sectional survey with a sample size N based on the relations derived in Section 2.2. Using a delta method expansion of the incidence estimator, we derive an approximate expression for its coefficient of variation. These expressions facilitate error estimation both for study design and data analysis. On calibration, we note that trends, as opposed to absolute values, for incidence can be obtained without any information about the distribution of finite waiting times under the threshold. However, an estimate of the fraction of assay non-progressors is essential. A key observation is that, for realistic population dynamics and sample sizes, statistical error is much larger than bias.

Numerical simulations are presented in Section 2.4. These demonstrate the application of the incidence estimator to a simulated population with epidemiological and demographic dynamics. This demonstrates reproducibility as well as bias introduced by imperfect calibration.

In the conclusion, we note that the framework presented here is quite general and is applicable to any TRI, as long as any non-zero probability of assay non-progression can be calibrated, survival is the same for assay progressors and non-progressors, and there is no regression below the recency threshold. It may be possible to modify the analysis to relax these requirements.

2.2 Relating the Prevalence of 'Recent Infection' to Incidence

We now outline a quite general approach to relating the key demographic, epidemiological and biological processes which are relevant to the estimation of incidence from cross-sectional surveys of the prevalence of 'recent infection'. This refines the naïve intuition that a high prevalence of 'recently infected' individuals means a high incidence.

2.2.1 The Basic Model

A test for recent infection, such as the CDC STARHS algorithm (Serological Testing Algorithm for Recent HIV Seroconversion [47]), is typically obtained through the administration of two assays of different sensitivity. The more sensitive test distinguishes infected from healthy individuals and the less sensitive test, applied to the infected individuals, distinguishes 'recent' from 'long' established infection.

Consider an assay which yields a quantitative result, the value of which typically increases with time from infection. The BED assay is of this type, the quantitative result being a normalized optical density (ODn), which is an increasing function of the proportion of HIV-1 specific IgG. Such an assay becomes the less sensitive component of a test for recent infection when we declare a threshold value and define 'observed to be recently infected' to be a test value under the threshold.

As there is inevitably inter-individual variation in the threshold crossing times, the category 'observed to be recently infected' is not sharply defined by a time boundary. We now adopt the more precise labels *under* threshold (U) and *over* threshold (O). The variability of times spent in the under-threshold category, conditional on being alive long enough to reach the threshold, is captured by a distribution of waiting times $f_{\mathrm{U|A}}$.

It is now possible to construct the basic epidemiological model shown in Figure 2.1A. Since our analysis will focus on a variety of *survival* functions $S(t)$, we shall refer to the susceptible population as the *healthy* population $H(t)$. Upon infection, individuals move from the healthy population to the under-threshold infected population $U(t)$. Those that live long enough reach the threshold after a waiting time, distributed according to $f_{\mathrm{U|A}}$, and enter the over-threshold population $O(t)$. We denote by $\tau_{\mathrm{U|A}}$ a waiting time generated by the density $f_{\mathrm{U|A}}$. The corresponding cumulative probability function is given by

$$F_{\mathrm{U|A}}(t) = \int_0^t f_{\mathrm{U|A}}(s)\, ds, \tag{2.1}$$

while the probability of 'survival' (persistence) in the population U, conditional on being alive, is

$$S_{\mathrm{U|A}}(t) = \mathbb{P}(\tau_{\mathrm{U|A}} > t) = 1 - F_{\mathrm{U|A}}(t), \tag{2.2}$$

and the mean waiting time is

$$\mathbb{E}\left[\tau_{\mathrm{U|A}}\right] = \int_0^\infty \tau f_{\mathrm{U|A}}(\tau)\, d\tau = \int_0^\infty S_{\mathrm{U|A}}(t)\, dt. \tag{2.3}$$

Analogously, we define f_{A}, τ_{A}, F_{A}, S_{A} and $\mathbb{E}\left[\tau_{\mathrm{A}}\right]$ to describe how long individuals remain *alive* after the moment of infection, and f_{U}, τ_{U}, F_{U}, S_{U} and $\mathbb{E}\left[\tau_{\mathrm{U}}\right]$ to describe how long individuals remain *simultaneously* alive and under the threshold on the

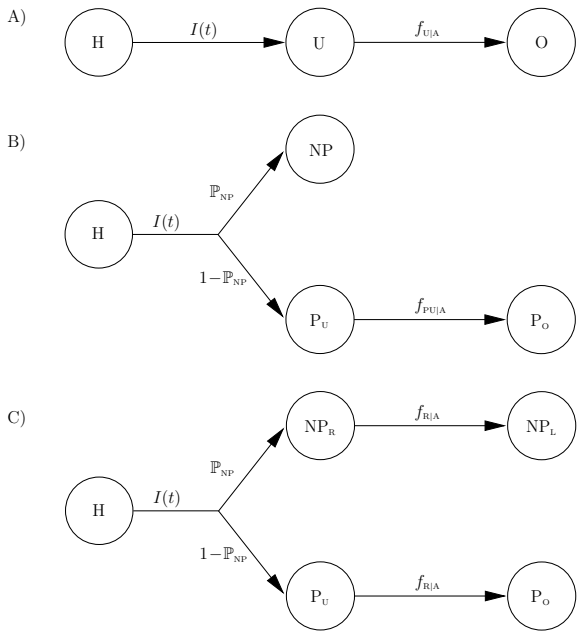


Fig. 2.1:

A) The basic epidemiological/TRI progression model. Members of the healthy population H are subject to a per unit time hazard (incidence) I of infection. After infection, individuals enter the under-threshold population U. Those that survive to progress into the over-threshold population O do so after delays distributed according to $f_{\mathrm{U|A}}$.

B) The basic model modified to accommodate non-progressors on the TRI. Now, upon infection, a proportion $\mathbb{P}_{\mathrm{NP}}$ of individuals remain under the threshold of the TRI for the rest of their lives, i.e., they enter the NP category. The remaining proportion, $1 - \mathbb{P}_{\mathrm{NP}}$, the progressors, enter the progressing, under-threshold population $\mathrm{P_U}$. Those that survive long enough enter the progressing, over-threshold population $\mathrm{P_O}$ with waiting times from $f_{\mathrm{PU|A}}$.

C) Modified model with separation of non-progressors into recent and long infected categories. This model contains the same epidemiology and biology as the model in B), with the introduction of a bookkeeping device which facilitates the definition of a calibratable category of recently infected individuals. The non-progressors are assigned waiting times drawn from the distribution observed in the progressing population, and spend this waiting time in the non-progressing recently infected ($\mathrm{NP_R}$) category, before moving to the non-progressing long infected ($\mathrm{NP_L}$) category.

assay. We assume that survival time and waiting time to threshold are independent in this model. Hence, the probability, at a time delay Δt after infection, of being *simultaneously* alive and under the threshold on the assay is

$$\mathbb{P}(\tau_{\mathrm{A}} > \Delta t \text{ AND } \tau_{\mathrm{U|A}} > \Delta t) = S_{\mathrm{A}}(\Delta t) S_{\mathrm{U|A}}(\Delta t) = S_{\mathrm{U}}(\Delta t). \tag{2.4}$$

Similarly, the probability of being *simultaneously* alive and over the threshold is

$$\mathbb{P}(\tau_{\mathrm{A}} > \Delta t \text{ AND } \tau_{\mathrm{U|A}} \leq \Delta t) = S_{\mathrm{A}}(\Delta t)(1 - S_{\mathrm{U|A}}(\Delta t)). \tag{2.5}$$

Hence, the mean time spent in the category U, accounting for both assay progression and mortality, is $\mathbb{E}[\tau_{\mathrm{U}}]$.

New infections are generated by a non-homogeneous Poisson process with an intensity (probability per unit time of new arrivals) $\lambda(t)$. Let the instantaneous incidence be given by $I(t)$. Then, in a period dt around time t, the expected number of new cases dC is given by

$$dC = \lambda(t)\, dt = I(t) H(t)\, dt. \tag{2.6}$$

We can now write down numerous expressions resulting from the model. For example, the expected number of historically accumulated cases, at time t, is given by

$$C(t) = \int_{-\infty}^{t} \lambda(s)\, ds = \int_{-\infty}^{t} I(s) H(s)\, ds. \tag{2.7}$$

The expected populations of infected persons under and over the threshold at time t are

$$\begin{aligned} U(t) &= \int_{-\infty}^{t} I(s) H(s) \mathbb{P}(\tau_{\mathrm{A}} > t - s \text{ AND } \tau_{\mathrm{U|A}} > t - s)\, ds \\ &= \int_{-\infty}^{t} I(s) H(s) S_{\mathrm{U}}(t - s)\, ds \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} O(t) &= \int_{-\infty}^{t} I(s) H(s) \mathbb{P}(\tau_{\mathrm{A}} > t - s \text{ AND } \tau_{\mathrm{U|A}} \leq t - s)\, ds \\ &= \int_{-\infty}^{t} I(s) H(s) S_{\mathrm{A}}(t - s)(1 - S_{\mathrm{U|A}}(t - s))\, ds. \end{aligned} \tag{2.9}$$

Our goal is to relate I for recent times to instantaneous values of H, U and O. Note that there is *fundamentally* a loss of information when one tries to characterize the history of a population based on observations made at a single time point. The recent historical course of a population, and even *instantaneous* values of state variables which are *rates*, like incidence, are in general not inferable from *counting* data

obtained in a single survey. This is due to the fact that instantaneous population states are, unavoidably, convolutions of historical epidemiological variables, as in (2.8) and (2.9) above. Any attempt to derive incidence estimates from the counting of infections accumulated in the recent past faces this problem, and at best some sort of weighted average of the recent values of incidence can be inferred without additional assumptions.

In general, a well defined construction of an estimate for incidence, based on data obtained in a survey conducted at time t, will be some sort of weighted average of past values

$$I_{\mathrm{W}}(t) = \frac{\int_{-\infty}^{t} I(s)W(s,t)\,ds}{\int_{-\infty}^{t} W(s,t)\,ds}, \tag{2.10}$$

where $W(s,t)$ is a statistical weight arising from a convolution of population history and biology. Since our goal is to estimate incidence from a count of recently infected individuals, a natural weighting function is one that reflects the relative contributions to this count made by infections from different times in the recent past. Hence, we consider

$$W(s,t) = H(s)S_{\mathrm{U}}(t-s) \tag{2.11}$$

since $W(s,t)$ is proportional to the probability that individuals are

1. available for being infected at time $s < t$, and
2. still alive and classified as under the threshold at time t, if infected at time s.

Using (2.11) as the weighting function leads to an expression for the weighted incidence given by

$$\begin{aligned} I_{\mathrm{W}}(t) &= \frac{\int_{-\infty}^{t} I(s)H(s)S_{\mathrm{U}}(t-s)\,ds}{\int_{-\infty}^{t} H(s)S_{\mathrm{U}}(t-s)\,ds} \\ &= \frac{U(t)}{\int_{-\infty}^{t} H(s)S_{\mathrm{U}}(t-s)\,ds}. \end{aligned} \tag{2.12}$$

The numerator in this expression is an instantaneous state variable, while the denominator in principle involves data from the entire history of the system as well as full knowledge of the survival function S_{U}.

A few remarks about the practical meaning of this weighted average are in order. In the case of constant incidence, the weighted average is the instantaneous value. In the case of a narrowly peaked distribution $f_{\mathrm{U|A}}$, a constant rate of change of I and a constant healthy population, the weighted average is approximately equal to the instantaneous incidence at a time $\mathbb{E}\left[\tau_{\mathrm{U}}\right]/2$ prior to the cross-sectional survey. If trends are fitted to the results of multiple cross-sectional surveys, this time lag could be more systematically accounted for.

2.2.2 A Simple Expression for Incidence

A simplified expression for weighted incidence in terms of sample and calibration data is now derived. We express the healthy population using the expansion

$$H(t+s) = H(t) + H_1 s + H_2 s^2 + \dots \tag{2.13}$$

and use the identity

$$\int_{-\infty}^{0} s^n S_{\mathrm{U}}(-s)\, ds = \tfrac{(-1)^n}{n+1} \mathbb{E}\left[\tau_{\mathrm{U}}^{n+1}\right], \tag{2.14}$$

which follows directly from integration by parts. It then follows that the weighted incidence (2.12) can be expressed as

$$\begin{aligned} I_{\mathrm{W}}(t) &= \frac{U(t)}{\int_{-\infty}^{t} H(s) S_{\mathrm{U}}(t-s)\, ds} \\ &= \frac{U(t)}{\int_{-\infty}^{0} H(t+s) S_{\mathrm{U}}(-s)\, ds} \\ &= \frac{U(t)}{H(t)\mathbb{E}\left[\tau_{\mathrm{U}}\right] - \frac{H_1}{2}\mathbb{E}\left[\tau_{\mathrm{U}}^2\right] + \dots}. \end{aligned} \tag{2.15}$$

If the healthy population is approximately constant for the times where the weight W is non-vanishing, we obtain the simple relation

$$I_{\mathrm{W}} \approx \frac{U(t)}{H(t)\,\mathbb{E}\left[\tau_{\mathrm{U}}\right]}, \tag{2.16}$$

which gives a weighted recent incidence in terms of instantaneous state variables (H and U) and the expected waiting time in the under-threshold category and is formally equivalent to the well known steady state result (for example [17]).

Expectation values of the form $\mathbb{E}\left[\tau_{\mathrm{U}}^n\right]$ are not state variables and should in principle be measured independently of a particular cross-sectional survey. Usually this would be accomplished in a calibration cohort follow-up study. Thus, after calibrating some of these expectation values, we can deal with a truncated expansion for H without further assumptions about the behavior of I.

It will be difficult to obtain accurate estimates of terms of higher order than just $\mathbb{E}\left[\tau_{\mathrm{U}}\right]$ for any assay forming the basis for a TRI. Finding the non-leading terms H_i (for $i \geq 1$) in the expansion of H will also require considerable demographic research. This means that one will most likely be constrained to use the simple expression (2.16), even if the healthy population is not approximately constant over the times where W is non-vanishing. The key question then is: how severe is the bias introduced by using the simple formula under realistic non-constancy of the healthy population?

Consider a non-constant healthy population given by $H(t) = H_0 e^{\alpha t}$. This has a conveniently tunable degree of failure to conform with the constancy assumption required for (2.16). When $\alpha = 0$ we have a constant number of healthy individuals, while a value of $\alpha = \ln(x)$ means the population grows by a factor of x in one year. We can provide a survival function $S_{\mathrm{U}}(t)$ for time measured in years, roughly inspired by the ODn progression on the BED assay, by specifying f_{U} to be a Weibull distribution with scale parameter $l = 0.44$ and shape parameter $k = 7$, corresponding to a mean of 150 days and a standard deviation of 25 days. We now numerically evaluate the denominator of (2.15) and compare it to the denominator of the simple formula (2.16). Note that this bias calculation is independent of the actual time dependence in I.

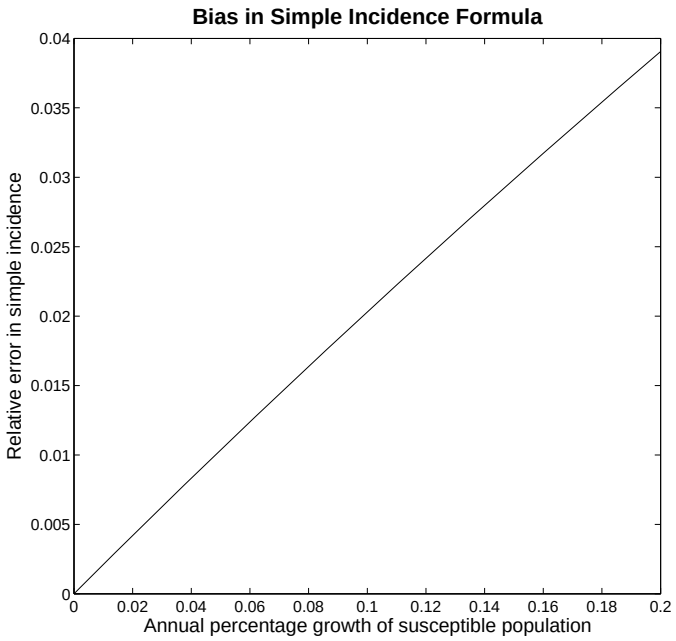


Fig. 2.2: Fractional error in the simple incidence relation (2.16) versus the full relation (2.15), as a function of growth rate of the healthy population, quoted as a percentage annual growth. The scenario is defined by: $H(t) = H(0)e^{\alpha t}$ and f_{U} is a Weibull distribution with scale parameter $l = 0.44$ and shape parameter $k = 7$. The parameter α is varied to produce deviation from a constant healthy population ($\alpha = 0$) in which limit equation (2.16) is exact.

In Figure 2.2, the bias in the simple formula (reported as a fraction of the unbiased value) is shown as a function of α, reported as the annual percentage growth. Note that, under the distributional assumptions made above, the bias is

about 2% for a population growing at an impressive rate of 10% per annum. As a rule of thumb, for a distribution of times spent under threshold which is not too diffuse around its mean, the bias, expressed as a percentage of the estimate, is approximately the annual percentage growth in the susceptible population multiplied by half the mean window period, measure in years. As we shall see later when analyzing a slightly more complex model of a TRI, this is a very minor source of error compared with the statistical error that arises as a result of using realistic sample sizes, not to mention imperfect calibration. Thus, in this example, bias arising from the non-constancy of the healthy population is not a key concern unless there is very dramatic variation in H, such as in a population of refugees. This analysis should in principle be carried out using the distributions applicable for any other assay intended for use as a TRI. The bias calculation demonstrated here is also applicable to the more complex model that now follows.

2.2.3 Modeling Assay Non-progressors

A known complicating factor for the BED assay (and likely also of any other TRI, such as an antibody avidity test) is that a small number of individuals utterly fail to progress beyond any practical threshold used to define recency. This is due to individual variation in biochemical details such as immune response, for example. The assay non-progression phenomenon leads to a long-term accumulation of *apparently* recently infected individuals, as classified by the TRI. The analysis in the previous section is perfectly valid even when assay non-progressors are present. However, if some individuals leave U only through death, possibly after long waiting times, then the calibration parameter $\mathbb{E}[\tau_{\mathrm{U}}]$ arises from a distribution with a long, difficult to characterize, tail. Moreover, the evolving context of real populations (eg. roll out of ART.) will presumably lead to these survival functions changing over time. Also, when the weighting (2.11) of the incidence estimator has support over longer periods, the estimator is less representative of recent incidence. We now extend the previous analysis in order to provide a better estimate of incidence accounting for the presence of assay non-progressors. Henceforth, we will use 'progression' to mean assay progression, not disease progression.

Consider the model captured in Figure 2.1B. At the moment of infection, individuals transition from the healthy population to either a non-progressing population (NP) or to a progressing under-threshold population ($\mathrm{P_U}$). The probability of non-progression is $\mathbb{P}_{\mathrm{NP}}$, and hence the probability of progression is $1 - \mathbb{P}_{\mathrm{NP}}$. Those individuals in $\mathrm{P_U}$ wait for a stochastic delay after which they move into the progressing over-threshold category $\mathrm{P_O}$ or die. In the previous model, $f_{\mathrm{U|A}}$ was the distribution of waiting times governing the transition, but since the waiting times

for non-progressing individuals are infinite, $f_{\mathrm{U|A}}$ can now not be normalized. Therefore, in order to specify the transition times from $\mathrm{P_U}$ to $\mathrm{P_O}$ in terms of a normalized density, we introduce the density of waiting times in the state of being a *progressor* and *under* the threshold, conditional on being a *alive*, denoted by $f_{\mathrm{PU|A}}$. Then $S_{\mathrm{U|A}}(t)$, $S_{\mathrm{PU|A}}(t)$ and $\mathbb{P}_{\mathrm{NP}}$ are related by

$$S_{\mathrm{U|A}}(t) = (1 - \mathbb{P}_{\mathrm{NP}})S_{\mathrm{PU|A}}(t) + \mathbb{P}_{\mathrm{NP}}. \tag{2.17}$$

The difficulty is that the TRI will classify as 'recently infected' all the individuals in the NP *and* $\mathrm{P_U}$ categories even though some potentially large number in NP are long infected. This can systematically be addressed by the following two key steps.

Firstly, we assume that the same survival function S_{A} is applicable to both progressing and non-progressing individuals. This is true if the differences between individuals which account for progression versus non-progression do not translate into significant differences in post-infection survival. This assumption has also been made, at least implicitly, in previous work on use of the BED assay for estimating HIV incidence (for example, see [117, 67] for analysis of [63]). Its applicability should in principle be tested for any assay used as a TRI.

Secondly, we introduce two artificial categories by separating the non-progressing population into 'recently infected' ($\mathrm{NP_R}$) and 'long infected' ($\mathrm{NP_L}$) sub-populations. Individuals entering the $\mathrm{NP_R}$ sub-population are assigned a waiting time drawn from $f_{\mathrm{PU|A}}$ after which they transition to the $\mathrm{NP_L}$ category. Note that this is a book-keeping device used for convenience and, unlike the assumption about survival, does not rely on any property of disease progression. It is now possible to provide a sensible definition for the class of *recently* infected individuals (R) which has a population given by

$$R(t) = P_{\mathrm{U}}(t) + NP_{\mathrm{R}}(t). \tag{2.18}$$

Note that, since both $\mathrm{P_U}$ and $\mathrm{NP_R}$ now have the same exit waiting times, the distribution of waiting times for R is given by $f_{\mathrm{R|A}} = f_{\mathrm{PU|A}}$, with corresponding survival function $S_{\mathrm{R|A}}$.

These two steps lead to the model in Figure 2.1C. It is now possible to recycle our preceding analysis and write down expected counts in these new classes. Survival in the state of being *simultaneously* alive and *recently infected*, is given by $S_{\mathrm{R}}(t) = S_{\mathrm{A}}(t)S_{\mathrm{R|A}}(t)$, and hence for the progressing populations we obtain

$$P_{\mathrm{U}}(t) = (1 - \mathbb{P}_{\mathrm{NP}}) \int_{-\infty}^{t} I(s)H(s)S_{\mathrm{R}}(t-s)\,ds \tag{2.19}$$

and

$$P_{\mathrm{O}}(t) = (1 - \mathbb{P}_{\mathrm{NP}}) \int_{-\infty}^{t} I(s)H(s)S_{\mathrm{A}}(t-s)(1 - S_{\mathrm{R|A}}(t-s))\,ds. \tag{2.20}$$

Note the similarity with expressions for $U(t)$ and $O(t)$ in the basic model. For the non-progressing populations we obtain

$$NP_{\mathrm{R}}(t) = \mathbb{P}_{\mathrm{NP}} \int_{-\infty}^{t} I(s)H(s)S_{\mathrm{R}}(t-s)\,ds \tag{2.21}$$

and

$$NP_{\mathrm{L}}(t) = \mathbb{P}_{\mathrm{NP}} \int_{-\infty}^{t} I(s)H(s)S_{\mathrm{A}}(t-s)(1-S_{\mathrm{R|A}}(t-s))\,ds. \tag{2.22}$$

For convenience we define

$$\varphi = \frac{\mathbb{P}_{\mathrm{NP}}}{1-\mathbb{P}_{\mathrm{NP}}}, \tag{2.23}$$

and note that

$$NP_{\mathrm{R}}(t) = \varphi P_{\mathrm{U}}(t) \tag{2.24}$$

and, more importantly,

$$NP_{\mathrm{L}}(t) = \varphi P_{\mathrm{O}}(t). \tag{2.25}$$

These equations express the symmetry between the progressing and non-progressing sub-populations of Figure 2.1C. Substituting (2.19) and (2.21) into (2.18), we can write

$$R(t) = \int_{-\infty}^{t} I(s)H(s)S_{\mathrm{R}}(t-s)\,ds. \tag{2.26}$$

It is appropriate to use a weighting scheme analogous to the one used in the basic model

$$W(s,t) = H(s)S_{\mathrm{R}}(t-s), \tag{2.27}$$

since $W(s,t)$ is now proportional to the probability that individuals are alive and classified as *recently infected* at time t if they become infected at time s, regardless of whether they are progressors or non-progressors. Then the weighted incidence, denoted I_{W}, is given by

$$\begin{aligned} I_{\mathrm{W}}(t) &= \frac{\int_{-\infty}^{t} I(s)H(s)S_{\mathrm{R}}(t-s)\,ds}{\int_{-\infty}^{t} H(s)S_{\mathrm{R}}(t-s)\,ds} \\ &= \frac{R(t)}{\int_{-\infty}^{t} H(s)S_{\mathrm{R}}(t-s)\,ds}. \end{aligned} \tag{2.28}$$

The populations of under-threshold and over-threshold individuals are related to the populations defined in Figure 2.1C by

$$U(t) = P_{\mathrm{U}}(t) + NP_{\mathrm{R}}(t) + NP_{\mathrm{L}}(t) \tag{2.29}$$

and

$$O(t) = P_{\mathrm{O}}(t). \tag{2.30}$$

Using the above two equations and (2.18) and (2.25), the population of recent infections is related to the under-threshold and over-threshold populations by

$$R(t) = U(t) - \varphi O(t). \tag{2.31}$$

Performing the same expansion as before and assuming a slowly varying healthy population gives

$$I_{\mathrm{W}}(t) \approx \frac{U(t) - \varphi O(t)}{H(t)\,\mathbb{E}\left[\tau_{\mathrm{R}}\right]}. \tag{2.32}$$

This expresses the incidence in terms of the calibration parameters $\mathbb{E}\left[\tau_{\mathrm{R}}\right]$ and φ (equivalently $\mathbb{P}_{\mathrm{NP}}$), and the state variables H, U and O.

All that has changed, as a result of allowing non-progressors into the model, is the shift in the numerator from U to $R = U - \varphi O$. The same bias calculations as before apply immediately, but there is an increase in statistical sensitivity.

2.3 Statistics and Calibration

The population models of the preceding section are expected to be in ever closer correspondence to a real population as the population size increases. To model the sampling process of a cross-sectional survey with a sample size N, we rescale the subpopulations of the continuous time model, at any time t, by the total population size $T = H + U + O$, to obtain the population proportions $\mathbb{P}_{\mathrm{H}} = H/T$, $\mathbb{P}_{\mathrm{U}} = U/T$ and $\mathbb{P}_{\mathrm{O}} = O/T$. The result of a survey employing the TRI is the set of three counts $N_{\mathrm{H}} + N_{\mathrm{U}} + N_{\mathrm{O}} = N$. We do not address the difficulties relating to study design and selection bias which would need to be confronted in the field, but proceed on the assumption that sampling is unbiased. For large populations, where the assumption of sampling with replacement is benign, the counts are trinomially distributed around their means. These observed counts turn equation (2.32) into an estimator for the recently weighted incidence I_{est} given by

$$I_{\mathrm{est}} = \frac{1}{\mathbb{E}\left[\tau_{\mathrm{R}}\right]} \frac{N_{\mathrm{U}} - \varphi N_{\mathrm{O}}}{N_{\mathrm{H}}}. \tag{2.33}$$

The quantities in this estimator can conveniently be regarded as being of two types:

1. population counts (N_{H}, N_{U} and N_{O}) observed in a cross-sectional survey, and
2. calibration parameters ($\mathbb{P}_{\mathrm{NP}}$ and $\mathbb{E}\left[\tau_{\mathrm{R}}\right]$) which are estimated from follow-up.

In the simplest application, each of these five variables is estimated just once. A delta method expansion (see, for example, Chapter 14 in [3]) leads to a coefficient of variation (CV) denoted by c_v:

$$c_v^2 = \frac{1}{N}\frac{1}{\mathbb{P}_{\mathrm{O}}+\mathbb{P}_{\mathrm{U}}}\left(\frac{1}{\mathbb{P}_{\mathrm{H}}}+\frac{\mathbb{P}_{\mathrm{O}}\mathbb{P}_{\mathrm{U}}(1+\bar{\varphi})^2}{(\mathbb{P}_{\mathrm{U}}-\bar{\varphi}\mathbb{P}_{\mathrm{O}})^2}\right)+\frac{\sigma_\omega^2}{\bar{\omega}^2}+\frac{\sigma_{\mathbb{P}_{\mathrm{NP}}}^2\mathbb{P}_{\mathrm{O}}^2}{(1-\bar{\mathbb{P}}_{\mathrm{NP}})^4(\mathbb{P}_{\mathrm{U}}-\bar{\varphi}\mathbb{P}_{\mathrm{O}})^2}. \quad (2.34)$$

In the appendix we define the conventions which lead to this expression. We note that the assumption of Gaussian uncertainty in the calibration parameters is a heuristic at best.

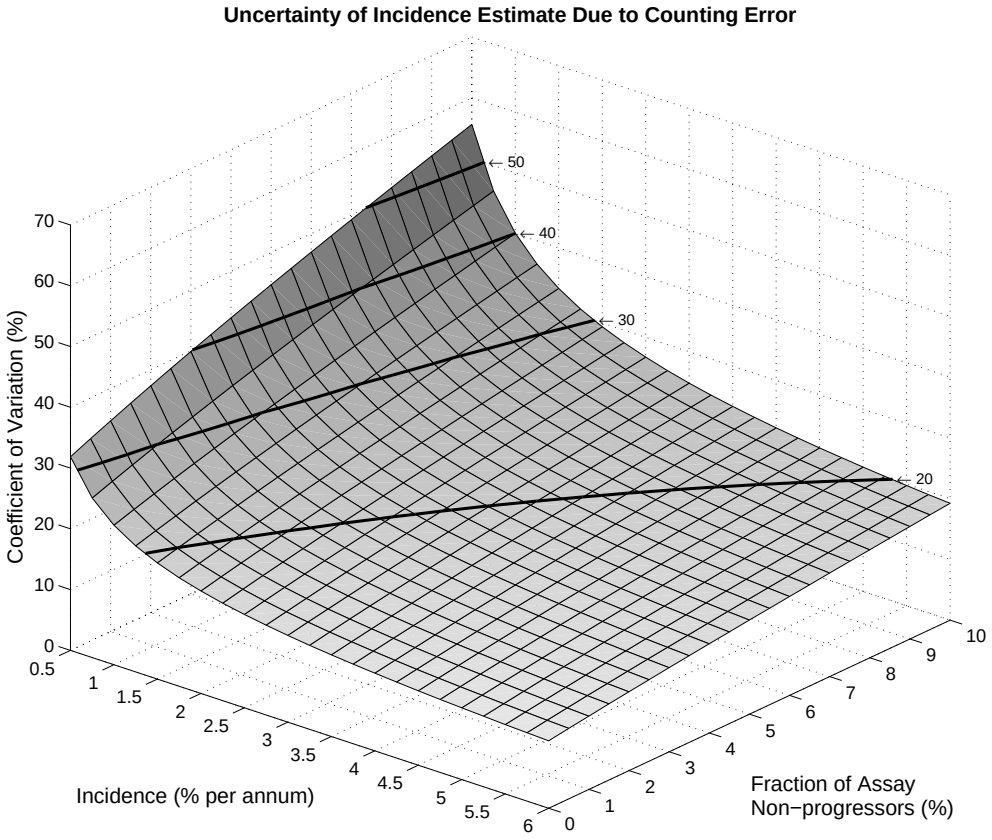


Fig. 2.3: Coefficient of variation (2.34) of the incidence estimator, under the assumption of perfect calibration (i.e. $\sigma_\omega = \sigma_{\mathbb{P}_{\mathrm{NP}}} = 0$). A sample size of 5000 is drawn from a steady state scenario implicit in choices for I and $\mathbb{P}_{\mathrm{NP}}$, given a mean post-infection survival of 8 years and a mean window period of 150 days. The bold lines are contours of constant CV.

Assuming perfect calibration and a sample size of 5000 individuals, Figure 2.3 shows how the counting error component of the CV depends on incidence and the fraction of assay-non-progressors for a steady state epidemic. The appendix outlines

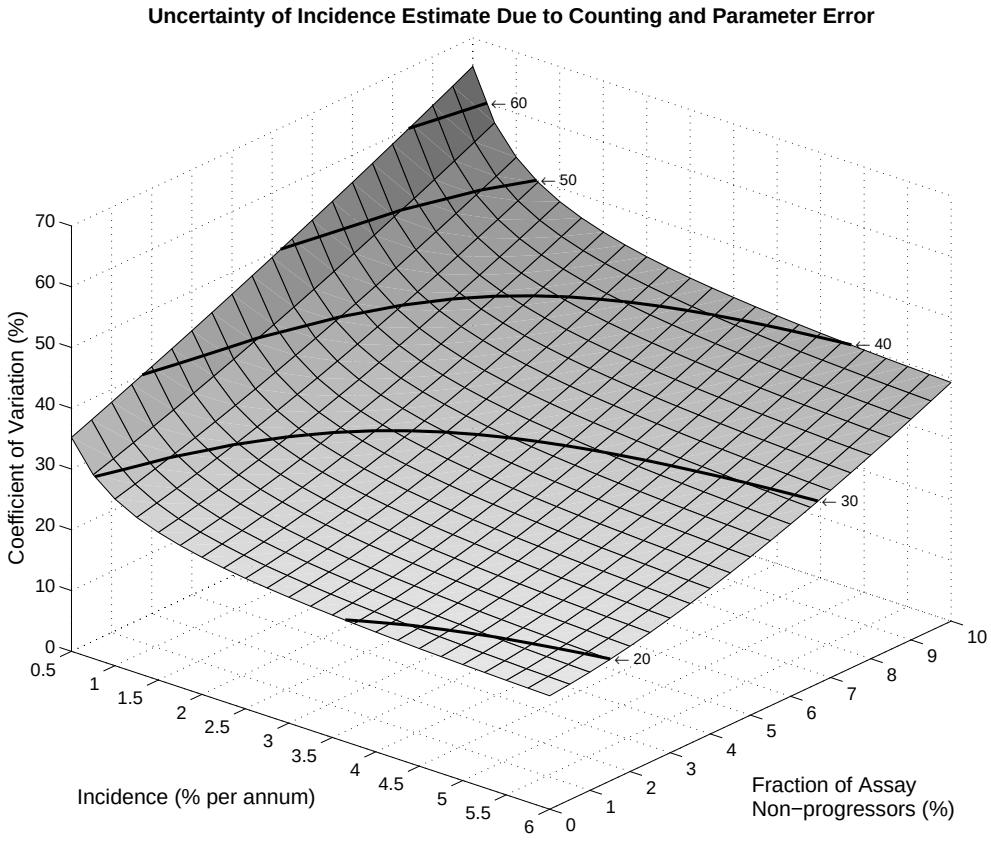


Fig. 2.4: Coefficient of variation (2.34) of the incidence estimator. A sample size of 5000 is drawn from a steady state scenario implicit in choices for I and $\mathbb{P}_{\text{NP}}$, given a mean post-infection survival of 8 years and a mean window period of 150 days. The parameters $\mathbb{P}_{\text{NP}}$ and ω are now assumed to be generated from normal distributions around their true values, with coefficients of variation of 15%. The bold lines are contours of constant CV.

how the steady state proportions were derived. Figure 2.4 shows the total CV for the same sample size when one assumes both calibration parameters are drawn from a normal distribution with a CV of 15%[2].

Although it is difficult to verify the accuracy of all terms in equation (2.34), it is straightforward to verify that it handles counting error with high precision. For a specific set of population proportions, a coefficient of variation can be computed using an enumeration of all possible trinomially distributed counts (excluding those

[2] A $\mathbb{P}_{\text{NP}}$ estimate of 5% obtained by following 850 seroconvertors would have a CV of 15%. The uncertainty associated with estimates of the mean window period depends in detail on methodological choices and is consequently more complicated to characterize.

that lead to an infinite estimate, i.e. when $N_{\mathrm{H}} = 0$). Comparing the coefficient of variation using the trinomial counts with the counting error component of (2.34), a maximum relative discrepancy of 0.01% was found for the range of values shown in Figure 2.3.

Depending on the study design, and the availability of locally relevant calibration data, various scenarios can be envisaged which involve more or less ability to reuse calibration parameters between cross-sectional surveys. For example, if one wishes merely to estimate *trends* in incidence, as opposed to absolute incidence values, then it is not necessary to have an estimate for $\mathbb{E}[\tau_{\mathrm{R}}]$ at all, since it is just an overall factor. However, surveys conducted at different times will not yield comparable values of $I_{\mathrm{est}} \propto (U - \varphi O)/H$ unless φ (equivalently $\mathbb{P}_{\mathrm{NP}}$) is known with some accuracy, since it appears in one of two terms in the numerator. Consider two surveys which use the same point estimate

$$\varphi = \bar{\varphi} + \delta_{\varphi}, \tag{2.35}$$

where $\bar{\varphi}$ is the real value and δ_{φ} is the error due to methodological and statistical factors. The first survey obtains counts $N_{\mathrm{H}}^{(1)}$, $N_{\mathrm{U}}^{(1)}$ and $N_{\mathrm{O}}^{(1)}$ and the second obtains counts $N_{\mathrm{H}}^{(2)}$, $N_{\mathrm{U}}^{(2)}$ and $N_{\mathrm{O}}^{(2)}$. This leads to an estimate of the difference between the two incidences of

$$\Delta I_{\mathrm{est}}(\varphi) = I_{\mathrm{est}}^{(1)}(\varphi) - I_{\mathrm{est}}^{(2)}(\varphi) = \frac{N_{\mathrm{U}}^{(1)} - \varphi N_{\mathrm{O}}^{(1)}}{N_{\mathrm{H}}^{(1)}\mathbb{E}[\tau_{\mathrm{R}}]} - \frac{N_{\mathrm{U}}^{(2)} - \varphi N_{\mathrm{O}}^{(2)}}{N_{\mathrm{H}}^{(2)}\mathbb{E}[\tau_{\mathrm{R}}]}. \tag{2.36}$$

Knowledge of the exact value $\bar{\varphi}$ leads to

$$\Delta I_{\mathrm{est}}(\bar{\varphi}) = I_{\mathrm{est}}^{(1)}(\bar{\varphi}) - I_{\mathrm{est}}^{(2)}(\bar{\varphi}) = \frac{N_{\mathrm{U}}^{(1)} - \bar{\varphi} N_{\mathrm{O}}^{(1)}}{N_{\mathrm{H}}^{(1)}\mathbb{E}[\tau_{\mathrm{R}}]} - \frac{N_{\mathrm{U}}^{(2)} - \bar{\varphi} N_{\mathrm{O}}^{(2)}}{N_{\mathrm{H}}^{(2)}\mathbb{E}[\tau_{\mathrm{R}}]}, \tag{2.37}$$

from which we see that the error in ΔI_{est}, due to the error in φ, is

$$\Delta I_{\mathrm{est}}(\varphi) - \Delta I_{\mathrm{est}}(\bar{\varphi}) = \frac{\delta_{\varphi}}{\mathbb{E}[\tau_{\mathrm{R}}]}\left(\frac{N_{\mathrm{O}}^{(2)}}{N_{\mathrm{H}}^{(2)}} - \frac{N_{\mathrm{O}}^{(1)}}{N_{\mathrm{H}}^{(1)}}\right). \tag{2.38}$$

The direction and magnitude of error depend in detail on many factors, such as population renewal and long-term post-infection survival. While it is not possible to summarize all the effects that may be produced by imperfect estimation of $\mathbb{P}_{\mathrm{NP}}$, in Section 2.4 we conduct a number of numerical simulations which demonstrate the kind of bias that may arise.

2.4 Numerical Simulations

In this section we present results of numerical simulations that demonstrate the use of the simple expression (2.33) for incidence estimation in a non-steady state epidemic.

Arrival times of new infections were generated according to a non-homogeneous Poisson process with intensity given by $\lambda(t) = H(t)I(t)$. Uniform random numbers (r_i) between 0 and 1 were generated, and new infection arrival times t_i computed as solutions of

$$\exp\left(-\int_{t_{i-1}}^{t_i} \lambda(t)\,dt\right) = r_i. \tag{2.39}$$

Newly infected individuals were initially classified as under the recency threshold of a TRI and assigned survival times generated by f_{A}. A fraction $1-\mathbb{P}_{\mathrm{NP}}$ progressed to the over-threshold category according to waiting times generated by $f_{\mathrm{R|A}}$. Weibull distributions were used for $f_{\mathrm{R|A}}$ and f_{A}, i.e. post-infection waiting times Δt solve

$$\exp\left(-\left(\frac{\Delta t}{l}\right)^k\right) = r, \tag{2.40}$$

where k and l are the relevant shape and scale parameters and r is a uniformly distributed random number between 0 and 1. Unique individuals were drawn from the population at intervals, to produce counts N_{H}, N_{U} and N_{O}, and hence estimates for incidence.

To demonstrate the incidence estimation process, a 50 year population scenario was produced. The Weibull shape and scale parameters for f_{A} and $f_{\mathrm{R|A}}$ were chosen to give approximately realistic values for the mean and standard deviations for the window period and infected life expectancy, as detailed in Table 2.1. The healthy population was set to $H(t) = 100,000 + 5,000t$, with t measured in years. The incidence was set at 0.01 (hazard per person per year) for the first ten years, climbing linearly to 0.1 over the next ten years, then remaining at this high level for a further ten years, followed by ten years of linear decline to 0.03 and maintained at this level for the last ten years of the simulation.

	Shape (k)	Scale (l)	Mean	Standard Dev.
Life expectancy (f_{A})	4.5	8.83	8 years	2 years
Window period ($f_{\mathrm{R|A}}$)	7	0.44	150 days	25 days

Tab. 2.1: Weibull parameters for the Monte Carlo simulation (survival time measured in years).

Figure 2.5 shows output from this simulation. The input incidence parameter is indicated as the dotted *instantaneous incidence* curve. A sample of 5,000 individuals was surveyed every year, and an incidence estimate was produced using the simple estimator (2.33) with exact values of $\mathbb{E}[\tau_{\mathrm{R}}]$ and $\mathbb{P}_{\mathrm{NP}}$, i.e., assuming perfect calibration. These point estimates are indicated as *estimated incidence* values, using '+' symbols. The combined effects of the previously noted time convolution in

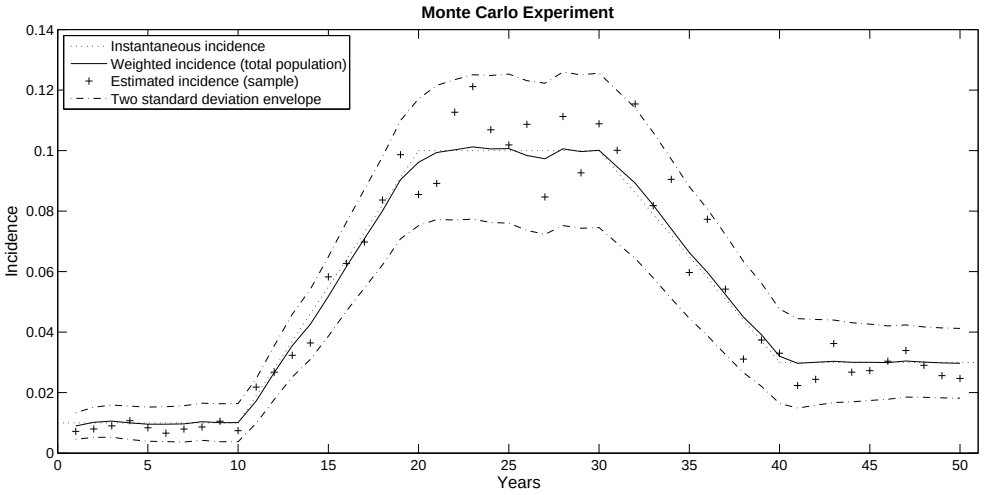


Fig. 2.5: Full stochastic simulation of population with epidemic, individual survival times, and annual sampling of $5,000$ individuals. The healthy population was set to $H(t) = 100,000 + 5,000t$ with time measured in years. The *instantaneous incidence* parameter is the dotted curve. The target of the estimates is the *weighted incidence* (solid line), which was calculated explicitly as per (2.28) from all the known inputs. This is flanked by a *two standard deviation counting error envelope* (dashed lines). Simulated *estimated incidence* (+ symbols) were obtained by using sample counts in the simple estimator (2.33). The calibration parameters $\mathbb{E}[\tau_{\mathrm{R}}]$ and $\mathbb{P}_{\mathrm{NP}}$ were assumed to be known exactly.

I_{W}, as well as stochastic departure from means in the simulated population, make the input incidence parameter an unrealistic target for simulated incidence measurements. Thus, the solid *weighted incidence* line has been displayed, which uses full knowledge of all population members' classification into H, U or O, inserted into (2.28) with full knowledge of the denominator, (both the non-constant $H(t)$ and the exact S_{R}). This is essentially all that the incidence estimation algorithm can be asked to reproduce. A *two standard deviation counting error envelope* around the *weighted incidence* line, calculated using the first term of (2.34) and knowledge of the full population state and calibration parameters, is shown as two dashed lines.

In Section 2.3 it was shown that incidence trends can be extracted without $\mathbb{E}[\tau_{\mathrm{R}}]$ calibration, while an estimate for $\mathbb{P}_{\mathrm{NP}}$ is vital. We now explore the extent to which the accuracy of the estimate of $\mathbb{P}_{\mathrm{NP}}$ affects the ability to determine a trend in incidence. Population fractions for H, U and O were extracted at six times from the population simulation described above and are shown in Table 2.2. Four instances of incidence trend estimation were simulated by selecting the time intervals $(15, 20)$,

$(20, 30)$, $(30, 35)$ and $(40, 50)$. We considered the trends that would be observed if incidence were measured at the beginning and end of each of these intervals. In order to focus on the bias introduced by imperfect estimation of $\mathbb{P}_{\text{NP}}$, rather than sample size effects, we assumed perfect knowledge of $\mathbb{P}_{\text{H}}$, $\mathbb{P}_{\text{U}}$ and $\mathbb{P}_{\text{O}}$. For each of these intervals, we calculated an incidence estimate at the beginning and end, as a function of the estimated value of $\mathbb{P}_{\text{NP}}$ (the true value being 0.05), assuming $\mathbb{E}[\tau_{\text{R}}]$ is known exactly. We also calculated the estimated fractional change in incidence, which does not depend on $\mathbb{E}[\tau_{\text{R}}]$. The results are shown in Figure 2.6, where the four intervals are referred to as scenarios A, B, C and D, respectively.

In each case, the effect of the error in the estimation of $\mathbb{P}_{\text{NP}}$ is quite different, as can be understood by considering how (2.38) is impacted by the system history. Note that case B and case D both simulate intervals over which incidence is approximately constant, but the impact (on the estimated incidence change) of incorrect estimation of $\mathbb{P}_{\text{NP}}$ does not even agree in sign. Negative incidence estimates are obtained in panels C and D above critical overestimates of $\mathbb{P}_{\text{NP}}$. In panel D this breakdown of the model results in the divergence of the fractional change in estimated incidence. In short, incorrect estimates of $\mathbb{P}_{\text{NP}}$ can lead to the fundamental breakdown of the inference scheme. This makes sense, as $\mathbb{P}_{\text{NP}}$ impacts the long-term accumulation of individuals in the P_{O} category.

	Time (years)					
	15	20	30	35	40	50
$\mathbb{P}_{\text{H}}$	0.850	0.688	0.576	0.602	0.694	0.814
$\mathbb{P}_{\text{U}}$	0.024	0.041	0.044	0.036	0.024	0.019
$\mathbb{P}_{\text{O}}$	0.126	0.271	0.380	0.362	0.282	0.167

Tab. 2.2: Population fractions in H, U and O within a 50 year epidemic scenario.

2.5 Discussion and Conclusion

We have presented a detailed analysis of relations between recent incidence in a population and counts of 'recently infected' individuals. These are in principle complex convolutions involving the epidemiological history as well as all information about the distribution of waiting times in the recently infected category. When the healthy population undergoes realistically modest variation on the time scale of the definition of recency implied by the TRI, we obtain simplified forms which incur very little bias. The simplified relations yield estimators which are shown to have considerably more variance than bias under realistic demographic and epidemiological

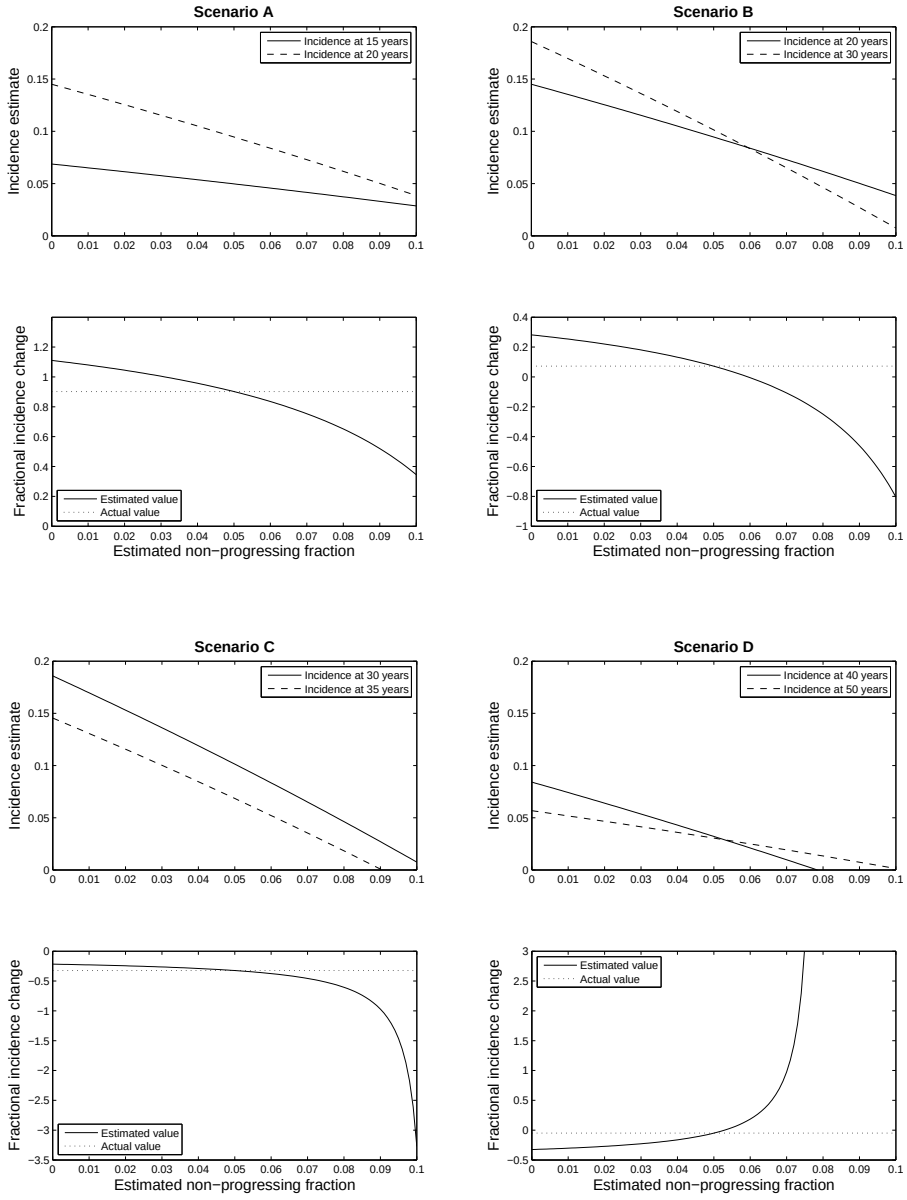


Fig. 2.6: Absolute incidence estimates and estimated fractional incidence changes for four pairs of successive times with population proportions from Table 2.2.

assumptions.

A key observation is that, for the purposes of estimating incidence from a TRI, there is no fundamental obstacle posed by having a *known* fraction of individuals fail to progress over the recency threshold. An *accurate* estimate of the non-progressing fraction alone, is sufficient (and necessary) to infer *trends* in incidence from cross-sectional surveys. However, as demonstrated in the calculations of Section 2.4, a suitably large error in the estimate of $\mathbb{P}_{\text{NP}}$ can render TRI based incidence estimates meaningless. This fraction could possibly be estimated for the BED assay from historical records, since there are many viable samples in storage with supporting clinical information indicating long-infected status. A calibration of the mean finite waiting time is required in order to estimate *absolute values* of incidence. Prospective follow-up is probably the only practical way to estimate this parameter.

In contrast to our model, which has only two calibration parameters, the well known model of McDougal et al. [63] appears to have four (window period, sensitivity, short-term specificity and long-term specificity). Hargrove et al. [40] have previously proposed a heuristic simplification of the McDougal approach, and we have recently shown that a rigorous simplification is possible under the original assumptions [117]. This allows a reduction of the parameters in the McDougal model to the ones that naturally occur in our approach. This has two advantages—our parameters are easier to calibrate, and assuming independence of correlated parameters leads to incorrect estimates of calibration error.

Noting that the assumptions of our model are the least restrictive of any TRI based incidence estimation method of which we are aware, we now consider its limitations. We have only modeled one direction of progression of individuals from an experimentally defined state of 'recent' infection to 'non-recent' infection. The reverse apparently occurs for BED optical density in some terminal stage AIDS patients and patients on ART. This process constitutes a substantial complication, and further work is required to investigate how it may be incorporated into an analytical model of the kind developed here. It may be worth exploring previous suggestions [63] to use additional information, from questionnaires or other assays, to remove end-stage/ART patients from the observed recent count. In accounting for assay non-progression we have assumed that the same post-infection survival applies to assay progressors and assay non-progressors. Data on the similarity or difference in mortality is preliminary at best, but we are aware of unpublished claims that the HIV long-term non-progressors are somewhat over represented in the population of BED assay non-progressors. We have also not considered the possibility that calibration parameters vary regionally, for example as a result of environmental impacts on immunity, or that they are functions of time, for example as a result

of substantial vaccine uptake in a population. All these complications are under investigation as part of ongoing work.

Besides the explicit assumptions noted, the analysis presented here is quite general. Tests for recent infection continue to be of interest, and new assays are likely to be developed both for HIV infection and other important diseases. In summary, we have presented a simple incidence estimator, which can inform design of appropriate calibration studies and cross-sectional incidence estimation surveys, and can also form the basis of systematic inference algorithms for processing the data obtained from such surveys. Our analysis provides a broad framework for a consistent approach to the estimation of non-constant hazard from instantaneous population counts, including explicit attention to the limits of validity and/or utility of such estimates in light of knowledge of the relevant survival functions.

2.6 Appendix

Given a sample of N subjects tested using the TRI, we use the delta method to derive a systematic error estimate for the estimator

$$I_{\text{est}} = \frac{N_{\text{U}} - \frac{\mathbb{P}_{\text{NP}}}{1-\mathbb{P}_{\text{NP}}} N_{\text{O}}}{\omega N_{\text{H}}}, \tag{2.41}$$

where $\omega = \mathbb{E}[\tau_{\text{R}}]$. The counts N_{X} fluctuate trinomially around their means $\bar{N_{\text{X}}} = \mathbb{P}_{\text{X}} N$. We assume the counts are sufficiently large so that binomial distributions can be approximated by normal distributions—which will be the case if the survey is to have any reasonable accuracy. In order to account for correlation, we express the three counts as the result of two independent random draws. We also assume that ω and $\mathbb{P}_{\text{NP}}$ fluctuate normally with standard deviations σ_ω and $\sigma_{\mathbb{P}_{\text{NP}}}$. The error estimate is derived using the delta method as follows:

- Let $\vec{\alpha} = [\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ be draws from a standard normal distribution.
- Set
$$N_{\text{H}} = \bar{N_{\text{H}}} + \sigma_{\text{H}} \alpha_1, \tag{2.42}$$
where
$$\sigma_{\text{H}} = \sqrt{N \mathbb{P}_{\text{H}} (1 - \mathbb{P}_{\text{H}})}. \tag{2.43}$$
- Set
$$N_{\text{U}} = \mathbb{P}_{\text{U}} N - \sigma_{\text{U}} \alpha_1 + \sigma_{\text{UO}}(\alpha_1) \alpha_2 \tag{2.44}$$
and
$$N_{\text{O}} = \mathbb{P}_{\text{O}} N - \sigma_{\text{O}} \alpha_1 - \sigma_{\text{UO}}(\alpha_1) \alpha_2 \tag{2.45}$$

where

$$\sigma_{\mathrm{U}} = \frac{\mathbb{P}_{\mathrm{U}}}{\mathbb{P}_{\mathrm{O}} + \mathbb{P}_{\mathrm{U}}}\sigma_{\mathrm{H}}, \qquad \sigma_{\mathrm{O}} = \frac{\mathbb{P}_{\mathrm{O}}}{\mathbb{P}_{\mathrm{O}} + \mathbb{P}_{\mathrm{U}}}\sigma_{\mathrm{H}} \tag{2.46}$$

and

$$\sigma_{\mathrm{UO}}(\alpha_1) = \frac{\sqrt{(N - \mathbb{P}_{\mathrm{H}}N - \sigma_{\mathrm{H}}\alpha_1)\mathbb{P}_{\mathrm{O}}\mathbb{P}_{\mathrm{U}}}}{\mathbb{P}_{\mathrm{O}} + \mathbb{P}_{\mathrm{U}}}. \tag{2.47}$$

- Set

$$\omega = \bar{\omega} + \sigma_{\omega}\alpha_3. \tag{2.48}$$

- Set

$$\mathbb{P}_{\mathrm{NP}} = \bar{\mathbb{P}}_{\mathrm{NP}} + \sigma_{\mathbb{P}_{\mathrm{NP}}}\alpha_4. \tag{2.49}$$

Substituting these expressions into the incidence estimator (2.41) and taking partial derivatives with respect to each α_i, we get

$$\left.\frac{\partial I_{\mathrm{est}}}{\partial \alpha_1}\right|_{\vec{\alpha}=0} = -\frac{\bar{I}_{\mathrm{est}}}{\sqrt{N(\mathbb{P}_{\mathrm{O}} + \mathbb{P}_{\mathrm{U}})\mathbb{P}_{\mathrm{H}}}} \tag{2.50}$$

$$\left.\frac{\partial I_{\mathrm{est}}}{\partial \alpha_2}\right|_{\vec{\alpha}=0} = \sqrt{\frac{\mathbb{P}_{\mathrm{U}}\mathbb{P}_{\mathrm{O}}}{N(\mathbb{P}_{\mathrm{O}} + \mathbb{P}_{\mathrm{U}})}}\frac{(1+\bar{\varphi})\bar{I}_{\mathrm{est}}}{\mathbb{P}_{\mathrm{U}} - \bar{\varphi}\mathbb{P}_{\mathrm{O}}} \tag{2.51}$$

$$\left.\frac{\partial I_{\mathrm{est}}}{\partial \alpha_3}\right|_{\vec{\alpha}=0} = -\frac{\sigma_{\omega}\bar{I}_{\mathrm{est}}}{\bar{\omega}} \tag{2.52}$$

$$\left.\frac{\partial I_{\mathrm{est}}}{\partial \alpha_4}\right|_{\vec{\alpha}=0} = -\frac{\sigma_{\mathbb{P}_{\mathrm{NP}}}\mathbb{P}_{\mathrm{O}}\bar{I}_{\mathrm{est}}}{(1 - \bar{\mathbb{P}}_{\mathrm{NP}})^2(\mathbb{P}_{\mathrm{U}} - \bar{\varphi}\mathbb{P}_{\mathrm{O}})}, \tag{2.53}$$

where

$$\bar{I}_{\mathrm{est}} = \frac{\mathbb{P}_{\mathrm{U}} - \bar{\varphi}\mathbb{P}_{\mathrm{O}}}{\bar{\omega}\mathbb{P}_{\mathrm{H}}} \qquad \text{and} \qquad \bar{\varphi} = \frac{\bar{\mathbb{P}}_{\mathrm{NP}}}{1 - \bar{\mathbb{P}}_{\mathrm{NP}}}. \tag{2.54}$$

The coefficient of variation ($c_v = \sigma(I_{\mathrm{est}})/\bar{I}_{\mathrm{est}}$) is thus given by

$$c_v^2 = \frac{1}{N}\frac{1}{\mathbb{P}_{\mathrm{O}} + \mathbb{P}_{\mathrm{U}}}\left(\frac{1}{\mathbb{P}_{\mathrm{H}}} + \frac{\mathbb{P}_{\mathrm{O}}\mathbb{P}_{\mathrm{U}}(1+\bar{\varphi})^2}{(\mathbb{P}_{\mathrm{U}} - \bar{\varphi}\mathbb{P}_{\mathrm{O}})^2}\right) + \frac{\sigma_{\bar{\omega}}^2}{\bar{\omega}^2} + \frac{\sigma_{\mathbb{P}_{\mathrm{NP}}}^2\mathbb{P}_{\mathrm{O}}^2}{(1 - \bar{\mathbb{P}}_{\mathrm{NP}})^4(\mathbb{P}_{\mathrm{U}} - \bar{\varphi}\mathbb{P}_{\mathrm{O}})^2}. \tag{2.55}$$

To evaluate this formula for a suitable range of inputs, we construct a family of steady state epidemics, tunable by varying I and $\mathbb{P}_{\mathrm{NP}}$ for fixed values of ω and the mean post-infection survival time Ω. At equilibrium, the ratio of recent infections to long infections is given by the ratio of times spent in these categories. Under the assumption that there is no mortality in the recent category, this can be written as

$$\frac{\mathbb{P}_{\mathrm{U}} - \varphi\mathbb{P}_{\mathrm{O}}}{\mathbb{P}_{\mathrm{O}} + \varphi\mathbb{P}_{\mathrm{O}}} = \frac{\omega}{\Omega - \omega}. \tag{2.56}$$

The equilibrium total prevalence is given by the product of the recruitment rate and mean post-infection survival:

$$\mathbb{P}_{\mathrm{O}} + \mathbb{P}_{\mathrm{U}} = I\mathbb{P}_{\mathrm{H}}\Omega. \tag{2.57}$$

The two equations above, together with $\mathbb{P}_{\mathrm{H}} + \mathbb{P}_{\mathrm{U}} + \mathbb{P}_{\mathrm{O}} = 1$, uniquely define the equilibrium proportions.

Chapter 3

A Comparison of Biomarker Based Incidence Estimators

∗ This chapter was coauthored with A. Welte [69].

Abstract

Background: Cross-sectional surveys utilizing biomarkers that test for recent infection provide a convenient and cost effective way to estimate HIV incidence. In particular, the BED assay has been developed for this purpose. Controversy surrounding the way in which false-positive results from the biomarker should be handled has lead to a number of different estimators that account for imperfect specificity. We compare the estimators proposed by McDougal et al., Hargrove et al. and McWalter & Welte.

Methodology/Principal Findings: The three estimators are analyzed and compared. An identity showing a relationship between the calibration parameters in the McDougal methodology is shown. When the three estimators are tested under a steady state epidemic, which includes individuals who fail to progress on the biomarker, only the McWalter/Welte method recovers an unbiased result.

Conclusions/Significance: Our analysis shows that the McDougal estimator can be reduced to a formula that only requires calibration of a mean window period and a long-term specificity. This allows simpler calibration techniques to be used and shows that all three estimators can be expressed using the same set of parameters. The McWalter/Welte method is applicable under the least restrictive assumptions and is the least prone to bias of the methods reviewed.

3.1 Introduction

Although prospective follow-up of an initially HIV-negative cohort is widely regarded as the "gold-standard" for estimating incidence, the idea of utilizing a biomarker to define a suitable class of "recently infected" individuals, and then to use the prevalence of this class as the basis for estimating HIV incidence, is attractive for a

number of reasons. Since this can be implemented using a cross-sectional survey, it is logistically simpler, cheaper and less prone to the biases that result from intervention and loss to follow-up.

The BED capture enzyme immunoassay (BED assay) has been developed for this purpose [79, 28] and widely used [8]. It measures the proportion of IgG that is HIV-1 specific as a normalized optical density (ODn). Since this proportion increases over time after the infection event, specifying an ODn threshold allows seropositive individuals to be classified as recently infected, if they are below threshold, and as non-recently infected, if they are above threshold. Initially, an incidence formula was proposed [79] that did not explicitly account for the possibility of assay non-progressors (i.e. individuals who never develop enough of an immunological response to cross the threshold). This method was similar to the earlier approaches of Brookmeyer and Quinn [17]], and Janssen et al. [47]. Later, the methodology proposed by McDougal et al. [63] was the first to deal with assay non-progressors. They derived an incidence formula which can be expressed in terms of the prevalence of below-threshold seropositive, above-threshold seropositive and seronegative individuals, and four assay calibration parameters, being the mean window period (ω), sensitivity (σ), short-term specificity (ρ_1) and long-term specificity (ρ_2). Introducing the long-term specificity parameter provided a way to quantify assay non-progression.

Two other incidence paradigms that explicitly account for assay non-progressors have since been formulated. Hargrove et al. [40] proposed a simpler incidence estimator which is equivalent to the McDougal estimator when one sets $\sigma = \rho_1$. Recently, we have also proposed a formally rigorous incidence paradigm [8], which accounts for assay non-progression using fewer assumptions than are made by McDougal et al. The parameters that emerge naturally in our estimator are a mean window period and a probability of not progressing on the assay (which can also be expressed as a long-term specificity).

A large portion of this paper is dedicated to an analysis of the assay parameters of the McDougal methodology, showing how they are related. By using a survival analysis formulation of the problem, we are able to write down precise expressions for the parameters. This allows us to derive a relationship between three of the parameters, which simplifies the McDougal estimator by showing that only ω and ρ_2, which are considerably easier to calibrate than σ and ρ_1, are required in the final formula. The reduction of the McDougal approach is important in that it shows that all three incidence estimators are, in effect, based on the same underlying parameters characterising the performance of the assay, and are therefore amenable to direct comparison.

We then compare the performance of the three incidence estimators by substi-

tuting analytic expressions for population counts, derived from a model steady state epidemic, into the various formulae. This analysis shows that only our formula [71] produces a bias-free result. Although the biases are typically small, we demonstrate, using numerical examples, that there are regimes where bias may be significant.

The paper is structured as follows: We start by describing the McDougal methodology and, in doing so, write down mathematical expressions for the assay calibration parameters. In the next section we restate the assumptions made by McDougal et al. in a mathematically precise manner. This allows us to derive the identity that shows the relationship between the parameters. We then present the three incidence formulae and compare them by inserting the population counts from a model steady state epidemic. We conclude the paper with a discussion of the implications of the identity and the steady state analysis.

3.2 The McDougal Methodology

Denote the number of individuals in a cross-sectional sample who are respectively *under-threshold*, *over-threshold* and *healthy* (susceptible) by N_{U}, N_{O} and N_{H}. Then the McDougal estimator [63] can be written as

$$\bar{I} = \frac{f N_{\mathrm{U}}}{f N_{\mathrm{U}} + \omega N_{\mathrm{H}}}, \tag{3.1}$$

where ω is specified in years and the "correction factor",

$$f := \frac{P_{\mathrm{t}}}{P_{\mathrm{o}}} = \frac{P_{\mathrm{o}} + \rho_2 - 1}{P_{\mathrm{o}}(\sigma - \rho_1 + 2\rho_2 - 1)}, \tag{3.2}$$

is the ratio of the "true" proportion P_{t} of recent infections and the proportion $P_{\mathrm{o}} = N_{\mathrm{U}}/(N_{\mathrm{U}} + N_{\mathrm{O}})$ of the HIV positive individuals that are under the threshold. This correction factor, which depends on subtle definitions for the sensitivity and specificity parameters, explicitly accounts for the fact that the BED assay imperfectly classifies individuals as "recently infected".

McDougal et al. calibrate these parameters using seroconversion panels which show BED optical density as a function of time since infection (some of these are published [79, 28]). The calibration occurs in two stages. A window period is estimated, and then estimates of the sensitivity, short-term specificity and long-term specificity are determined with respect to the window period.

The window period is estimated as "the mean period of time from initial seroconversion to reaching an ODn of 0.8" [63]. Although it is not explicitly stated, we presume that those individuals that never reach the threshold, either because they do not progress above the threshold or because they die before reaching the threshold, are not included in the calculation of the mean. More specifically this implies

that the window period is the mean observable threshold crossing time, conditional on assay progression (i.e. actually reaching the threshold).

In order to calibrate the sensitivity, short-term specificity and long-term specificity, "a plot of the proportion of specimens positive in the assay versus time since seroconversion" is generated (also later referred to as "the curve"). This is the sampled survival function (essentially a Kaplan-Meier curve) in the state of being under the threshold, conditional on being alive, which we denote $S_{\mathrm{U|A}}(t)$.

The sensitivity of the test is estimated for an interval corresponding to the window period by "integrating the curve within the window". Short-term specificity is calculated for "the interval immediately after, and equal in duration to, the window period". Long-term specificity is for "the period thereafter (where the curve is flat)". McDougal et al. explicitly make the following assumptions, with the justification that they "are reasonable as very little attrition (from death) during the first two time intervals after infection would be expected":

1. "Recent infections are randomly distributed within the first window period".
2. "The number of persons in the interval of equal duration immediately after the mean window period equals the number in the first window period".
3. "The remainder of the population is more than two window periods since seroconversion".

While it may be true in the situation being explored here, we note that it is not *a priori* obvious that the choice of equal window periods ensures that $S_{\mathrm{U|A}}(t)$ is flat after twice the window period. With this in mind, we propose a generalization in which there are two window periods with arbitrary values ω_1 and ω_2, chosen so that all individuals that progress do so in a time less than $\omega_1 + \omega_2$ after seroconversion (i.e. $S_{\mathrm{U|A}}(t)$ is flat for $t > \omega_1 + \omega_2$, see the bottom graph of Figure 3.1). It should be noted that this is a special survival curve in that it never reaches a zero value, capturing the fact that a certain proportion of individuals will never progress above the threshold. This is what differentiates this approach from other approaches that do not account for assay non-progression (Such as Brookmeyer and Quinn [17], Janssen et al. [47], and Parekh et al. [79]).

For analytical convenience, we introduce $S_{\mathrm{PU|A}}(t)$, the survival of assay *progressors* in the state of being *under-threshold*. We also introduce $\mathbb{P}_{\mathrm{NP}}$, the probability of individuals *not progressing* on the assay. Then $S_{\mathrm{U|A}}(t)$, $S_{\mathrm{PU|A}}(t)$ and $\mathbb{P}_{\mathrm{NP}}$ are related by

$$S_{\mathrm{U|A}}(t) = (1 - \mathbb{P}_{\mathrm{NP}})S_{\mathrm{PU|A}}(t) + \mathbb{P}_{\mathrm{NP}}.$$

The introduction of $S_{\mathrm{PU|A}}(t)$ allows us to provide a precise definition of the window period used by McDougal et al. It is the mean time between seroconversion

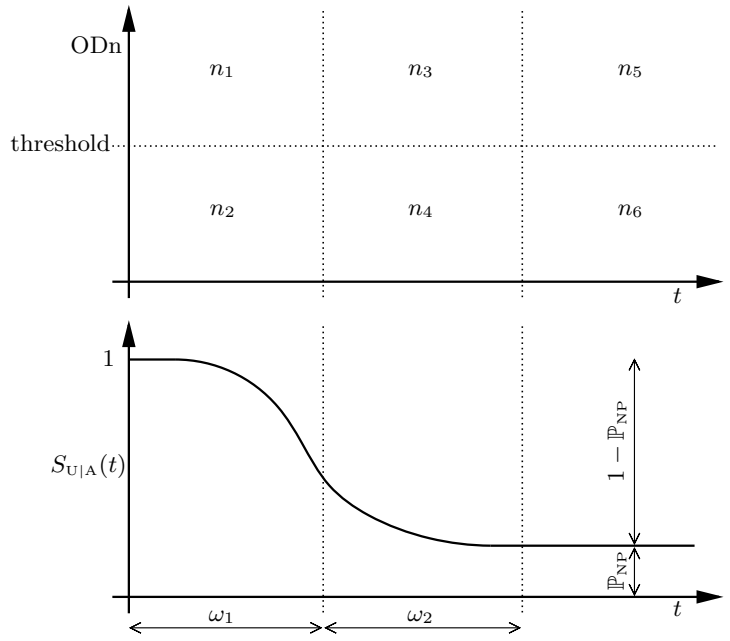


Fig. 3.1: The six sector model of McDougal et al. The top graph shows counts n_i and the bottom graph shows the survival functions $S_{\mathrm{U|A}}(t)$ versus time since infection.

and reaching threshold, for individuals who progress:

$$\omega = \int_0^\infty S_{\mathrm{PU|A}}(t)\,dt. \tag{3.3}$$

Assumption 1 above can only mean that infection times in the first window period are *uniformly distributed*. Although Assumption 2 merely states that the number of infections in the second window period is equal to the number in the first, we shall see later that for ρ_1 to be a property of the assay, independent of the epidemic state, we require the stronger assumption that the infection events in the second window period are also *uniformly distributed* with the same intensity as in the first window period. We see below that this assumption is implicit in the work of McDougal et al. To make this more explicit, we define $f(t)$ to be the density of times since infection realized in the sample. The number of seropositive individuals is then given by

$$N_{\mathrm{sp}} = \sum_{i=1}^{6} n_i = \int_0^\infty f(t)\,dt,$$

where n_i are the counts of individuals in the various categories depicted in the top graph in Figure 3.1.

Setting $f(t) = f_0$ over the first two window periods means that the ratio of the

number of infected individuals in the second window period to those in the first period is ω_2/ω_1. Assumption 2 is recovered when the length of the window periods is equal. It should be noted that depends on incidence, susceptible population and life expectancies over the history of the epidemic. With reference to Figure 3.1, we are now in a position to write expressions for the number of seropositive individuals in each sector:

$$
\begin{aligned}
n_1 &= \int_0^{\omega_1} f(t)(1 - S_{\mathrm{U|A}}(t))\,dt = f_0(1-\mathbb{P}_{\mathrm{NP}})\int_0^{\omega_1}(1-S_{\mathrm{PU|A}}(t))\,dt \\
n_2 &= \int_0^{\omega_1} f(t)S_{\mathrm{U|A}}(t)\,dt = f_0\omega_1\mathbb{P}_{\mathrm{NP}} + f_0(1-\mathbb{P}_{\mathrm{NP}})\int_0^{\omega_1} S_{\mathrm{PU|A}}(t)\,dt \\
n_3 &= \int_{\omega_1}^{\omega_1+\omega_2} f(t)(1 - S_{\mathrm{U|A}}(t))\,dt = f_0(1-\mathbb{P}_{\mathrm{NP}})\int_{\omega_1}^{\omega_1+\omega_2}(1-S_{\mathrm{PU|A}}(t))\,dt \\
n_4 &= \int_{\omega_1}^{\omega_1+\omega_2} f(t)S_{\mathrm{U|A}}(t)\,dt = f_0\omega_2\mathbb{P}_{\mathrm{NP}} + f_0(1-\mathbb{P}_{\mathrm{NP}})\int_{\omega_1}^{\omega_1+\omega_2} S_{\mathrm{PU|A}}(t)\,dt \\
n_5 &= \int_{\omega_1+\omega_2}^{\infty} f(t)(1 - S_{\mathrm{U|A}}(t))\,dt = (1-\mathbb{P}_{\mathrm{NP}})\int_{\omega_1+\omega_2}^{\infty} f(t)\,dt \\
n_6 &= \int_{\omega_1+\omega_2}^{\infty} f(t)S_{\mathrm{U|A}}(t)\,dt = \mathbb{P}_{\mathrm{NP}}\int_{\omega_1+\omega_2}^{\infty} f(t)\,dt.
\end{aligned}
$$

Using the above expressions, the sensitivity, the short-term specificity and the long-term specificity are given by

$$
\begin{aligned}
\sigma &= \frac{n_2}{n_1+n_2} = \frac{(1-\mathbb{P}_{\mathrm{NP}})\int_0^{\omega_1} S_{\mathrm{PU|A}}(t)\,dt + \omega_1\mathbb{P}_{\mathrm{NP}}}{\omega_1} \\
\rho_1 &= \frac{n_3}{n_3+n_4} = \frac{(1-\mathbb{P}_{\mathrm{NP}})\int_{\omega_1}^{\omega_1+\omega_2}(1-S_{\mathrm{PU|A}}(t))\,dt}{\omega_2} \\
\rho_2 &= \frac{n_5}{n_5+n_6} = 1-\mathbb{P}_{\mathrm{NP}}.
\end{aligned}
$$

We can now see why the assumption of uniformly distributed infection events for the first and second window periods is required—it is the only way in which a cancelation of $f(t)$ in the expressions for σ and ρ_1 is possible. Note that under bias-free recruitment into a survey, at time $t = 0$, we have

$$
f(t) = \frac{N_{\mathrm{sp}}}{T_{\mathrm{sp}}} I(-t)H(-t)S_{\mathrm{A}}(t), \tag{3.4}
$$

where $I(t)$ is the instantaneous incidence,$H(t)$ is the number of healthy (susceptible) individuals, $S_{\mathrm{A}}(t)$is the life-expectancy survival function measured from the time since infection and

$$
T_{\mathrm{sp}} = \int_0^{\infty} I(-t)H(-t)S_{\mathrm{A}}(t)\,dt
$$

is the total number of seropositive individuals alive in the population at the time of the survey. The ratio $N_{\mathrm{sp}}/T_{\mathrm{sp}}$ is just the fraction of the total population that has

been recruited. Thus, the only sensible way to ensure that $f(t) = f_0$, for $t < \omega_1 + \omega_2$, is to assume that the incidence and the susceptible population are constant, and the survival function $S_{\mathrm{A}}(t) = S_{\mathrm{A}}(0) = 1$.

We also see why $S_{\mathrm{U|A}}(t)$ must be flat after both window periods—this ensures that $S_{\mathrm{U|A}}(t)$ is constant and can be pulled out of the integrals in the expressions for n_5 and n_6 as the factor $\mathbb{P}_{\mathrm{NP}}$. This is necessary for ρ_2 to be independent of $f(t)$.

Furthermore, in order to specify ρ_2 so that it is independent of the state of the epidemic, an implicit assumption is being made that survival is the same for assay progressors and assay non-progressors. Note that $f(t)$ appears in the expressions for both n_5 and n_6. If different life expectancies were used in these formulae, reflecting a difference in survival for assay progressors and assay non-progressors, the f's in these formulae would need to be different, and would not cancel in the expression for ρ_2. This assumption is not explicitly stated by McDougal et al. but is implicit in their assumption that ρ_2 is independent of epidemic state.

With the calibration parameters specified in the more general setting of unequal window periods ω_1 and ω_2, we now generalize the expression for the correction factor

$$f = \frac{P_{\mathrm{t}}}{P_{\mathrm{o}}},$$

where $P_{\mathrm{t}} = (n_1 + n_2)/N_{\mathrm{sp}}$ is the proportion of seropositive individuals who are *truly* infected at a time less than ω_1. Recalling that $P_{\mathrm{o}} = (n_2 + n_4 + n_6)/N_{\mathrm{sp}}$ and using the definitions of the parameters, it is easy to verify that

$$P_{\mathrm{o}} = P_{\mathrm{t}}\sigma + P_{\mathrm{t}}\frac{\omega_2}{\omega_1}(1-\rho_1) + \left(1 - P_{\mathrm{t}} - P_{\mathrm{t}}\frac{\omega_2}{\omega_1}\right)(1-\rho_2).$$

This means that the correction factor can be expressed as

$$f = \frac{P_{\mathrm{o}} + \rho_1 - 1}{P_{\mathrm{o}}\left[\sigma - \frac{\omega_2}{\omega_1}\rho_1 + \left(1 + \frac{\omega_2}{\omega_1}\right)\rho_2 - 1\right]}. \tag{3.5}$$

Note that this equation simplifies to the previous expression (3.2) when one sets $\omega_1 = \omega_2$.

3.3 Elimination of Parameters

For completeness, we now provide a precise specification of the assumptions that are required in order to facilitate the analysis in the rest of this paper. We note that with the exception of arbitrary sized window periods, these assumptions are equivalent to the assumptions—either explicit or implicit—that are being made by McDougal et al. [63].

Model Assumptions. *Specify window periods ω_1 and ω_2. We assume that:*

1. *The window periods are chosen so that the survival function $S_{\mathrm{U|A}}(t)$ is flat (and equal to $\mathbb{P}_{NP}$) for $t > \omega_1 + \omega_2$. This means that $S_{\mathrm{PU|A}}(t)$ only has support on the time interval $t \in [0, \omega_1 + \omega_2]$.*

2. *Arrival times of infection events are uniformly distributed on the interval $[0, \omega_1 + \omega_2]$. An equivalent way of stating this assumption is that over the interval $t \in [0, \omega_1 + \omega_2]$, $H(t)$ and $I(t)$ are constant and $S_{\mathrm{A}}(t) = 1$.*

3. *Survival is the same for assay progressors and assay non-progressors.*

We are now able to provide the identity relating the parameters in the McDougal approach.

Proposition 3.1. *Under the model assumptions stated above, the following identity holds:*

$$\sigma - \frac{\omega_2}{\omega_1}\rho_1 + \left(1 + \frac{\omega_2}{\omega_1} - \frac{\omega}{\omega_1}\right)\rho_2 = 1. \tag{3.6}$$

Proof. Since we assume that $S_{\mathrm{PU|A}}(t)$ only has support on $t \in [0, \omega_1 + \omega_2]$, we have

$$\int_0^{\omega_1+\omega_2} S_{\mathrm{PU|A}}(t)\,dt = \int_0^{\infty} S_{\mathrm{PU|A}}(t)\,dt = \omega.$$

Then, simply evaluating

$$\begin{aligned}
\sigma - \frac{\omega_2}{\omega_1}\rho_1 &= \frac{(1-\mathbb{P}_{\mathrm{NP}})\int_0^{\omega_1} S_{\mathrm{PU|A}}(t)\,dt + \omega_1\mathbb{P}_{\mathrm{NP}}}{\omega_1} \\
&\quad - \frac{\omega_2}{\omega_1}\frac{(1-\mathbb{P}_{\mathrm{NP}})\int_{\omega_1}^{\omega_1+\omega_2}(1 - S_{\mathrm{PU|A}}(t))\,dt}{\omega_2} \\
&= \frac{(1-\mathbb{P}_{\mathrm{NP}})\int_0^{\omega_1+\omega_2} S_{\mathrm{PU|A}}(t)\,dt - \int_{\omega_1}^{\omega_1+\omega_2}(1-\mathbb{P}_{\mathrm{NP}})\,dt + \omega_1\mathbb{P}_{\mathrm{NP}}}{\omega_1} \\
&= \frac{(1-\mathbb{P}_{\mathrm{NP}})(\omega - \omega_2 - \omega_1) + \omega_1}{\omega_1} \\
&= 1 - \left(1 + \frac{\omega_2}{\omega_1} - \frac{\omega}{\omega_1}\right)\rho_2,
\end{aligned}$$

yields the result directly. ∎

Using the proposition, the correction factor (3.5) simplifies to

$$f = \frac{\omega_1}{\omega}\frac{P_{\mathrm{o}} + \rho_2 - 1}{P_{\mathrm{o}}\rho_2}.$$

This expression no longer relies on estimates for σ and ρ_1. It is also interesting to note that it does not depend explicitly on ω_2. Calibrating ρ_2, however, requires

identifying individuals who have been infected for at least $\omega_1 + \omega_2$. Thus, ω_2 need not be precisely known, but a safe upper bound for $\omega_1 + \omega_2$ is required.

Furthermore, if we set $\omega_1 = \omega$ as in McDougal et al. then we recover

$$f = \frac{P_\circ + \rho_2 - 1}{P_\circ \rho_2}. \tag{3.7}$$

Note that (3.2) as stated in McDougal et al. contains three calibration parameters (σ, ρ_1 and ρ_2), while (3.7) contains only one calibration parameter (ρ_2). Incidence estimates using (3.1) and (3.7), however, still require the estimation of ω. The method of McDougal et al. can in principle be applied to an arbitrarily declared (as opposed to measured) window period, as long as σ, ρ_1 and ρ_2 are calibrated for that value. We have therefore reduced the number of calibration parameters by one.

Estimation of extra parameters may unnecessarily dilute the statistical power of the calibration data at hand. Moreover, estimates of the uncertainty due to calibration, based on the assumption of the independence of σ, ρ_1 and ρ_2, will be incorrect. Note that when one sets $\omega_1 = \omega_2 = \omega$, the identity is reduced to

$$\sigma - \rho_1 + \rho_2 = 1.$$

Substituting the estimates of the parameters found by McDougal et al., namely $\sigma = 0.768$, $\rho_1 = 0.723$ and $\rho_2 = 0.944$, into this equation gives a value of $0.989 \approx 1$ for the left hand side. The slight discrepancy is a manifestation of the combined fluctuations in the estimates of σ, ρ_1, ρ_2 and ω. Although ω is superficially absent in the identity, it enters as the period over which the other parameters are defined.

When one assumes that $\rho_2 = 1$ (corresponding to the situation where there are no assay non-progressors) and $\omega_1 = \omega$, the identity reduces to

$$\omega(1 - \sigma) = \omega_2(1 - \rho_1). \tag{3.8}$$

and the ratio of counts over this period is given by

$$\frac{n_1 + n_2}{n_3 + n_4} = \frac{\omega}{\omega_2}.$$

Using this ratio and substituting the definitions for σ and ρ_1 into (3.8), yields $n_1 = n_4$. Therefore, for tests with perfect long-term specificity, the observed count of individuals who are under-threshold is an unbiased estimate of the number of infections in the last period ω. This was noted in a less general analysis of Brookmeyer [15] where assay non-progressors were *a priori* excluded.

It should be noted that there is a subtlety in the definition of the window period that emerges in the above analysis. If, instead of (3.3), the window period is defined by

$$\omega := \int_0^\infty S_{\mathrm{PU|A}}(t) S_{\mathrm{A}}(t)\, dt. \tag{3.9}$$

then the two definitions are equivalent under the model assumptions leading to the proposition. This follows from the fact that $S_{\mathrm{PU|A}}(t)$ only has support on $t \in [0, \omega_1 + \omega_2]$ and that $S_{\mathrm{A}}(t) = 1$ over that interval. We have suggested an alternative incidence estimation paradigm [71] which requires fewer assumptions than the method of McDougal et al. In this approach $\mathbb{P}_{\mathrm{NP}}$ and ω, as defined in (3.9), emerge as the natural calibration parameters.

3.4 Comparison of Estimators Under Steady State Conditions

We now provide a simplified form for the McDougal incidence estimator based on the proposition. Substituting the new correction factor (3.7) into their estimator (3.1) and expressing the result in terms N_{U}, N_{O} and N_{H} gives

$$\bar{I}_a = \frac{N_{\mathrm{U}} - \mathbb{P}_{\mathrm{NP}}(N_{\mathrm{U}} + N_{\mathrm{O}})}{N_{\mathrm{U}} - \mathbb{P}_{\mathrm{NP}}(N_{\mathrm{U}} + N_{\mathrm{O}}) + \omega(1 - \mathbb{P}_{\mathrm{NP}})N_{\mathrm{H}}}. \tag{3.10}$$

where ω is specified in years. Here the subscript a indicates that the estimator is quoted as an "annualized incidence". Note that in writing down this expression, we have chosen to use $\mathbb{P}_{\mathrm{NP}}$ rather than the long-term specificity as this is a biologically more intuitive parameter. In addition, the other two estimators to which this estimator will be compared were originally specified in terms of $\mathbb{P}_{\mathrm{NP}}$.

In a previous attempt to simplify the McDougal formula, Hargrove et al. [40] proposed the following incidence formula

$$\tilde{I}_a = \frac{N_{\mathrm{U}} - \mathbb{P}_{\mathrm{NP}}(N_{\mathrm{U}} + N_{\mathrm{O}})}{N_{\mathrm{U}} - \mathbb{P}_{\mathrm{NP}}(N_{\mathrm{U}} + N_{\mathrm{O}} + N_{\mathrm{H}}) + \omega N_{\mathrm{H}}}. \tag{3.11}$$

where ω is specified in years. Note that they use the symbol ε where we use $\mathbb{P}_{\mathrm{NP}}$.

We have recently rigorously derived a weighted incidence estimator under less restrictive assumptions than those that are required for the McDougal or Hargrove approach [71]. Unlike the other two estimators, our estimator is expressed as a rate (indicated by a subscript r) and is given by

$$\begin{aligned} \hat{I}_r &= \frac{N_{\mathrm{U}} - \frac{\mathbb{P}_{\mathrm{NP}}}{1-\mathbb{P}_{\mathrm{NP}}} N_{\mathrm{O}}}{\omega N_{\mathrm{H}}} \\ &= \frac{N_{\mathrm{U}} - \mathbb{P}_{\mathrm{NP}}(N_{\mathrm{U}} + N_{\mathrm{O}})}{\omega(1 - \mathbb{P}_{\mathrm{NP}})N_{\mathrm{H}}}. \end{aligned} \tag{3.12}$$

To convert between an annualized incidence and an incidence expressed as a rate, one can use the standard conversion formula

$$I_a = 1 - e^{-I_r T} \qquad \Leftrightarrow \qquad I_r = \frac{-\ln(1 - I_a)}{T},$$

where $T = 1$ year.

In Appendix 3.6 we show that, under steady state conditions, N_{U} and N_{O} are specified in terms of N_{H} and an incidence rate I as

$$N_{\mathrm{O}} = IN_{\mathrm{H}}(1 - \mathbb{P}_{\mathrm{NP}})(\alpha - \omega) \tag{3.13}$$

$$\tag{3.14}$$

and

$$N_{\mathrm{U}} = IN_{\mathrm{H}}(1 - \mathbb{P}_{\mathrm{NP}})\omega + IN_{\mathrm{H}}\mathbb{P}_{\mathrm{NP}}\alpha \tag{3.15}$$

where α is the post-infection life expectancy. Using these population counts, it is now possible to compare the performance of the incidence estimators. Substituting (3.14) and (3.15) into the McDougal formula (3.10) yields

$$\bar{I}_a = \frac{I}{I+1}.$$

Converting this to a rate, we have

$$\bar{I}_r = \ln(I+1) = I + O(I^2),$$

where the last step results from a Taylor series expansion. Thus the estimator is accurate for small values of I , but yields a discrepancy at $O(I^2)$. The reason for this discrepancy is subtle. In deriving the correction factor, McDougal et al. assume uniform infection events over the window periods. We have shown that this is consistent with assuming that the incidence and susceptible population are constant. In using this factor to estimate an incidence with (3.1) they have, however, inconsistently assumed that these infection events are generated in a susceptible population which is being depleted by the infection events over a period of a year. This is implied by their choice of denominator in that formula, which adds back an annualized number of recent infections into the susceptible population. This is at odds with the assumption of a constant susceptible population, and leads to dimensionally inconsistent incidence estimators, (3.1) and (3.10).

To illustrate the magnitude of the bias, Figure 3.2 shows the difference between the McDougal incidence estimate and the equilibrium incidence, expressed as a percentage. Note that the range of incidence values used is large (up to 50% per annum). Although incidence for HIV is not likely to be larger than about 15% in the highest risk groups (e.g. injection drug users [44]), if this methodology were used to monitor other rapidly spreading epidemics, where incidence is large when stated in units of years, it would certainly produce unacceptable bias.

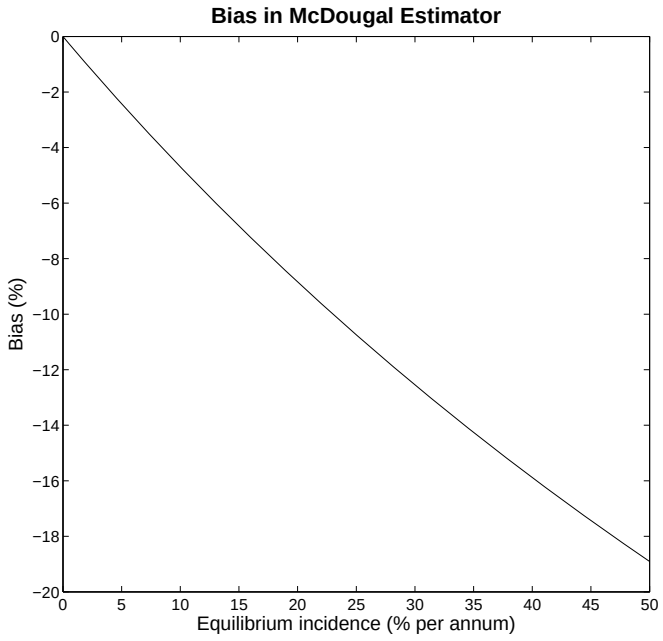


Fig. 3.2: Bias in the McDougal estimator. Relative difference between the McDougal estimate and the equilibrium incidence plotted as a function of equilibrium incidence.

Substituting the counts into the Hargrove formula (3.11) yields

$$\tilde{I}_a = \frac{I}{I + \frac{\omega - \mathbb{P}_{\mathrm{NP}}}{(1-\mathbb{P}_{\mathrm{NP}})\omega}},$$

which, when converted to a rate, gives

$$\tilde{I}_r = \ln\left(\frac{I(1-\mathbb{P}_{\mathrm{NP}})\omega}{\omega - \mathbb{P}_{\mathrm{NP}}} + 1\right) = I + \frac{\mathbb{P}_{\mathrm{NP}}(1-\omega)}{\omega - \mathbb{P}_{\mathrm{NP}}} I + O(I^2).$$

The Hargrove estimator incorporates the same form of denominator which leads to the second order discrepancy and dimensional inconsistency in the McDougal formula, and, in addition, it includes a linear bias term. Figure 3.3 demonstrates the bias introduced as a function of ω and $\mathbb{P}_{\mathrm{NP}}$ for an equilibrium incidence of 5% per annum. Although the bias is worst in the regimes where all the estimators have little statistical power and are unlikely to be used, there are nevertheless intermediate regimes where the bias is significant. Note that the estimator produces the same result (and bias) as the McDougal estimator when $\mathbb{P}_{\mathrm{NP}}$ or $\omega = 1$.

Finally, substituting the counts into our formula (3.12), which is already specified as a rate, yields

$$\hat{I}_r = I.$$

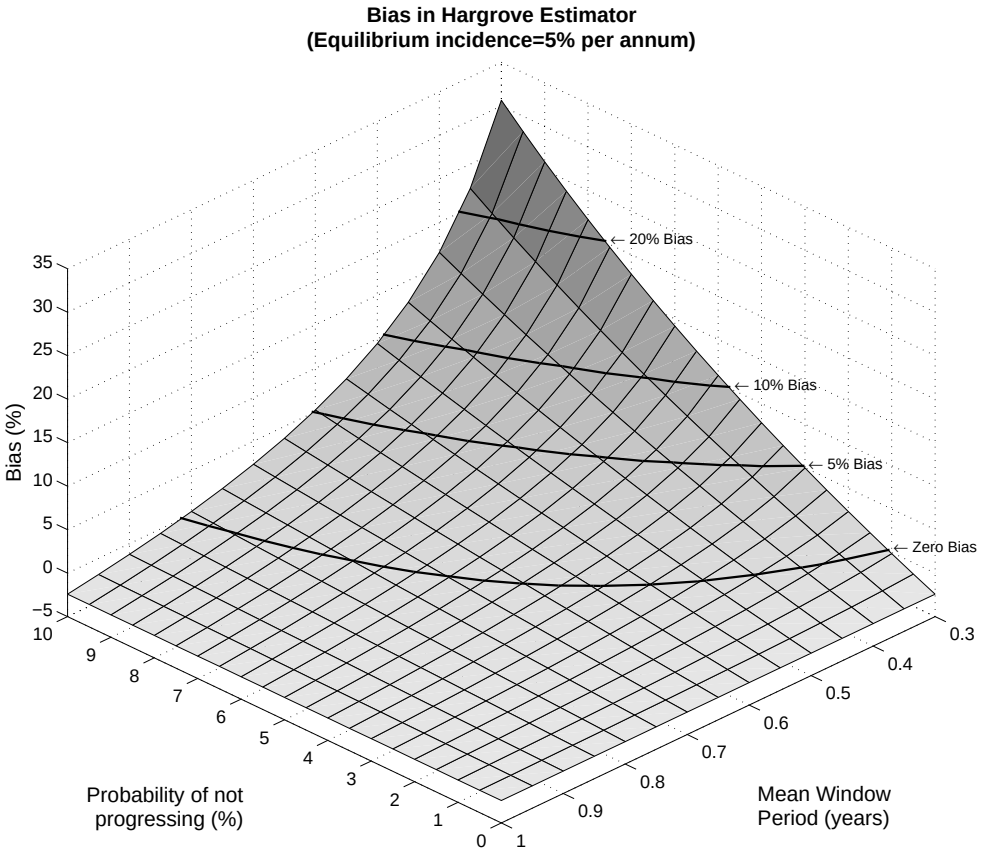


Fig. 3.3: Bias in the Hargrove estimator. Relative difference between the Hargrove estimate and the equilibrium incidence plotted as a function of ω and $\mathbb{P}_{\mathrm{NP}}$ for an equilibrium incidence of 5% per annum. Black lines indicate contours of equal bias.

Thus, under the assumption of a steady state epidemic, our weighted incidence estimator recovers the steady state incidence exactly. It is also the maximum likelihood estimator. This can be seen by writing the estimator in terms of the population proportions

$$\hat{I}_r = \frac{\mathbb{P}_{\mathrm{U}} - \frac{\mathbb{P}_{\mathrm{NP}}}{1-\mathbb{P}_{\mathrm{NP}}}\mathbb{P}_{\mathrm{O}}}{\omega \mathbb{P}_{\mathrm{H}}}, \qquad \text{where} \qquad \mathbb{P}_{\mathrm{X}} = \frac{N_{\mathrm{X}}}{N_{\mathrm{sp}}},$$

and noting that, since the counts are trinomially distributed, the sample proportions are the maximum likelihood estimates of the population proportions. We have already seen that the estimator solves for the equilibrium incidence. Thus, by the invariance property of maximum likelihood estimators (see e.g. p. 105 of van den Bos [106]), it is the maximum likelihood estimator for the incidence. This has also recently been demonstrated by Wang and Lagakos [112] by explicit maximization of

the log likelihood function.

A weighted incidence will in general not be equal to the instantaneous incidence under non-steady state conditions. We should, however, demand that any incidence formula exactly recover the incidence under this rather idealized situation.

3.5 Discussion

We have shown that under a precise restatement of the McDougal et al. assumptions, there exists a redundancy in the parameters they chose to characterise the assay. This allows the elimination of σ and ρ_1 from their estimator, with the important advantage that the remaining parameters are easier to calibrate. The calibration of σ and ρ_1 requires obtaining specimens from individuals with confidence about their time since infection (i.e. using frequent follow-up). On the other hand both ω and $\mathbb{P}_{\text{NP}}$ (or equivalently ρ_2) can be estimated through long follow-up intervals. The estimate for $\mathbb{P}_{\text{NP}}$ is the proportion of under-threshold samples known to be obtained more than $\omega_1+\omega_2$ post-infection. Given an estimate for $\mathbb{P}_{\text{NP}}$, an estimate of ω can be obtained from data with follow-up intervals greater than $\omega_1+\omega_2$ using an extended version [54] of the Bayesian approach previously described by Welte [115].

We have also shown that under steady state conditions the only estimator that is dimensionally consistent and produces an unbiased result is the one we have previously derived [71]. It is also the maximum likelihood estimator. The new approach makes fewer assumptions than the other methods. In particular, it consistently accounts for a dynamic epidemic by adopting a weighted definition of incidence. This overcomes a drawback of the other two methods which assume epidemic equilibrium for at least a period equal to the maximum progression time ($\omega_1+\omega_2$). It should be noted that this methodology is applicable to any biomarker, not only the BED assay—all that is needed is a suitable calibration of the assay parameters. It also follows that cross-sectional incidence estimates using this approach are applicable to infections other than HIV, as long as suitably calibrated assays that test for recent infection are available.

A shortcoming of all the methods explored here is that they make the assumption, either implicitly or explicitly, that survival for assay non-progressors and assay progressors is the same. As we have shown, relaxing this assumption means that the long-term specificity becomes epidemic state dependent and hence is time dependent. We are involved in ongoing work to address this issue.

3.6 Appendix

We derive the population level counts associated with a steady state epidemic. Assume that the number of susceptible individuals and the incidence are constant, and that the sample for our incidence calculation consists of the entire population. Let the susceptible population be $H(t) = N_{\mathrm{H}}$ and the incidence, expressed as a rate, be $I(t) = I$. Since our sample is the whole population we have $N_{\mathrm{sp}} = T_{\mathrm{sp}}$. Then, from (3.4) we obtain

$$f(t) = IN_{\mathrm{H}}S_{\mathrm{A}}(t),$$

and the number of over-threshold and under-threshold individuals in the total population are given by

$$\begin{aligned} N_{\mathrm{O}} &= n_1 + n_2 + n_3 \\ &= \int_0^\infty f(t)(1 - S_{\mathrm{U|A}}(t))\, dt \\ &= IN_{\mathrm{H}}(1 - \mathbb{P}_{\mathrm{NP}}) \int_0^\infty S_{\mathrm{A}}(t) - S_{\mathrm{PU|A}}(t)S_{\mathrm{A}}(t)\, dt \\ &= IN_{\mathrm{H}}(1 - \mathbb{P}_{\mathrm{NP}})(\alpha - \omega) \end{aligned}$$

and

$$\begin{aligned} N_{\mathrm{U}} &= n_2 + n_4 + n_6 \\ &= \int_0^\infty f(t)S_{\mathrm{U|A}}(t)\, dt \\ &= IN_{\mathrm{H}} \int_0^\infty (1 - \mathbb{P}_{\mathrm{NP}})S_{\mathrm{PU|A}}(t)S_{\mathrm{A}}(t) + \mathbb{P}_{\mathrm{NP}}S_{\mathrm{A}}(t)\, dt \\ &= IN_{\mathrm{H}}(1 - \mathbb{P}_{\mathrm{NP}})\omega + IN_{\mathrm{H}}\mathbb{P}_{\mathrm{NP}}\alpha, \end{aligned}$$

where α is the post infection life expectancy. It must be stressed that the survival functions $S_{\mathrm{A}}(t)$ and $S_{\mathrm{PU|A}}(t)$ are arbitrary. Thus, apart from assuming constant incidence and susceptible population, this is a quite general model.

Chapter 4

A Simplified Formula for Inferring HIV Incidence from Cross-Sectional Surveys Using a Test for Recent Infection

∗ This chapter was coauthored with A. Welte and T. Bärnighausen [117], and is reproduced with permission from Mary Ann Liebert, Inc: Aids Research and Human Retroviruses (2009) 25:125-6 DOI: 10.1089/aid.2008.0150.

4.1 Correspondence

The paper of McDougal et al. [63] is becoming a standard reference used for the estimation of HIV incidence from applications of the BED IgG-Capture Enzyme Immunoassay (BED assay) to cross-sectional blood samples [48, 86]. Their approach provides an estimate for an annual risk of infection in a hypothetical cohort, using an estimate for the true proportion, P_{t}, of 'recent infections' amongst HIV-seropositive individuals. The estimate P_{t} is in turn derived from the proportion, P_{o}, of seropositive individuals in a survey who test below a threshold value for normalized BED optical density (OD-n) [79]. The condition of being below the OD-n threshold is declared to be an imperfect test for recent infection.

True 'recent infection' is defined as having been infected for less than a period ω, where ω is the mean time individuals spend below the OD-n threshold. Since it is well known that not all individuals progress to a given threshold, even after arbitrarily long times, ω needs to be carefully defined as the mean threshold crossing time amongst those who *do* progress. It is also known that during late stage illness, or under the influence of antiretroviral therapy, individuals may regress to OD-n values below the recency threshold. It is further plausible, and indeed appears to be the case [10, 52], that the parameters characterising progression through the BED-defined states of infection vary regionally. These complications have caused doubt about the prospects for using the BED assay as a robustly characterisable test for recent infection for the purposes of estimating HIV incidence, as reflected in

a UNAIDS statement in 2006 [105] recommending it not be used for this purpose.

Hence, new assays, or combinations of assays (such as a BED and an antibody avidity test) are being developed, to provide more robust tests for recent infection. The fraction of individuals who progress atypically through an assay-defined class of 'recently infected' may thus be reduced, but is unlikely to be zero. Therefore, the methodology developed to deal with this problem for the BED assay appears, at face value, to be immediately transferable, requiring only minor modification (namely in the values of its parameters) to be applicable to other imperfect tests for recent infection. We argue that several subtle points need to be addressed to ensure that incidence inferences based on imperfect tests for recent infection are not unnecessarily limited, or even in error, and we do this by a critique of the original application.

The inter-individual variability of BED OD-n progression is captured in the McDougal model by three parameters:

- The sensitivity (σ) of the BED assay as a test for the condition of being 'recently infected', as defined above.
- The short term specificity (ρ_1) of the BED assay as a test for the condition of being 'recently infected', when restricted to persons who have been infected for a time between ω and 2ω.
- The long term specificity (ρ_2) of the BED assay as a test for the condition of being 'recently infected', when restricted to persons who have been infected for a time longer than 2ω.

Using data from a major epidemiological and demographic surveillance study in South Africa [99, 9], we and our collaborators are currently comparing various approaches to HIV incidence estimation using the BED assay [10, 12]. Given the long intervals between follow-up visits in this study (about a year), it was not possible to calibrate the McDougal formula in its published form. Calibration of σ and ρ_1 requires a follow-up interval of at most ω (which is of the order of half a year [63]).

While trying to address this issue, we discovered that a simplification of the McDougal formula is possible. In their paper, the key result relating P_{t} to the calibration parameters is given by

$$P_{\mathrm{t}} = \frac{P_{\mathrm{o}} + \rho_2 - 1}{\sigma - \rho_1 + 2\rho_2 - 1}. \tag{4.1}$$

As is shown by McWalter and Welte in a separate short note [67], the above equation can be simplified using the following identity:

$$\sigma - \rho_1 + \rho_2 = 1. \tag{4.2}$$

This identity is derived using no more assumptions than are used by McDougal et al. to derive their formula—these assumptions are, however, stated with greater precision in [67]. The idea that these parameters might be related was inspired by the analysis of the incidence estimation problem undertaken in [68]. Inserting the identity into (4.1) gives

$$P_{\mathrm{t}} = \frac{P_{\mathrm{o}} + \rho_2 - 1}{\rho_2}. \tag{4.3}$$

This means that, in order to estimate incidence, one needs only to calibrate the long term specificity ρ_2 (to estimate P_{t}) and the window period ω (to convert P_{t} to an annual risk of infection). Unlike σ and ρ_1, these can both be inferred from infrequent follow-up. Incidentally, using the values of σ, ρ_1, and ρ_2 reported in [63], we find that

$$\sigma - \rho_1 + \rho_2 = 0.989 \approx 1, \tag{4.4}$$

which manifests the combined fluctuations in the estimates of σ, ρ_1, ρ_2 and ω. Although ω is superficially absent in the identity, it enters as the period over which the other parameters are defined.

The appropriately simplified form (4.3) is amenable to calibration using data obtained with long intervals of follow-up [12]. This seems to us to be an important point, as many demographic and epidemiological surveillance studies we are aware of, or expect to see implemented, are characterized by follow-up intervals of the order of a year—almost ideal for calibrating the reduced formula, and clearly inadequate for calibrating the previously published form. There is likely to be substantial data of this sort available. On the other hand, the cost of obtaining short interval follow-up data is high, and the opportunities for doing so are rare.

Note that even given an appropriate data set for estimating σ, ρ_1, and ρ_2, the use of the naive formula, for the purpose of systematically quantifying uncertainty due to imperfect calibration, would require additional specification of non-trivial covariances implied by the identity (4.2).

The attraction of using a test for recent infection for HIV surveillance, programme evaluation and policy making, lies in the fact that it allows HIV incidence estimation from cross-sectional blood samples. Cross-sectional HIV status information alone, however, does not allow estimation of the calibration parameters. These must be estimated in separate studies, involving follow-up of an intensity comparable to a prospective observation of incidence. Only once this has been done can the more efficient cross sectional survey be employed on a suitably similar population. The more robust and locally validated the calibration parameters are, the more informative cross sectional surveys can be. Therefore it is important that the necessary parameters be calibrated as widely and thoroughly as possible, using such data

as is available. The parameters of the simplified formula are independent and can be estimated from comparatively long interval follow-up data, while the parameters used by McDougal et al. have non-trivial correlation and require short intervals of follow-up.

Chapter 5

Using Tests for Recent Infection to Estimate Incidence: Problems and Prospects for HIV

∗ This chapter was coauthored with A. Welte, O. Laeyendecker and T.B. Hallet [118].

Abstract

Tests for recent infection (TRIs), such as the BED assay, provide a convenient way to estimate HIV incidence rates from cross-sectional survey data. Controversy has arisen over how the imperfect performance of a TRI should be characterised and taken into account. Recent theoretical work is providing a unified framework within which to work with a variety of TRI- and epidemic-specific assumptions in order to estimate incidence using imperfect TRIs, but suggests that larger survey sample sizes will be required than previously thought. This paper reviews the framework qualitatively and provides examples of estimator performance, identifying the characteristics required by a TRI to estimate incidence reliably that should guide the future development of TRIs.

5.1 Introduction

When monitoring HIV epidemics it is vital to estimate incidence in order to plan and evaluate HIV programs [60]. Prospective cohort studies are the most direct way to achieve this. They are, however, expensive, prone to recruitment and retention bias, and potentially rendered unrepresentative by ethical obligations. The use of prevalence data in conjunction with mathematical modelling is an alternative approach [38, 97], but is indirect and requires accurate knowledge of mortality and migration. The disadvantages of these methods have focused attention on estimating incidence from cross-sectional surveys [17, 47, 80, 63, 40], with the result that a number of assays and algorithms that test for recent infection have been developed [64, 74]. In the context of HIV, such an assay or algorithm has sometimes been termed a STARHS (Serologic Testing Algorithm for Recent HIV Seroconversion) [64, 74], but we prefer to use the generic term 'test for recent infection' (TRI), because it does not specify a particular disease and method of testing. Recently,

the World Health Organization (WHO) Technical Working Group on Statistical Approaches for Development, Validation and Use of HIV Incidence Assays has proposed using the term 'recent infection testing algorithm' (RITA). The term has not, however, gained universal acceptance.

TRIs identify HIV-positive individuals who have been infected recently. By using a TRI in a serosurvey, incidence (I) can be estimated by applying the epidemiological relationship[1]:

$$I = \frac{R}{SD},$$

where R and S are the counts of 'recently infected' and 'susceptible' (HIV-uninfected) individuals observed in the cross-sectional survey and D is the mean duration spent in the 'recently infected' state, often called the (mean) window period. This incidence estimate is an average of the instantaneous incidence over a period of approximately D prior to the survey. The problem of incidence estimation then reduces to measuring the prevalence of 'recent infection', given knowledge of its duration.

TRIs usually discriminate recent from established infections by measuring specific aspects of the immune system which evolve during the course of initial infection. For HIV, this is typically the antibody response, with the titre, proportion of HIV-specific IgG, or antibody avidity (or a combination of these) providing quantitative output [74]. Laboratory defined thresholds are chosen to convert these outputs into categorical results. These results may be augmented with other clinical information, such as CD4 lymphocyte counts and antiretroviral therapy (ART) status, to classify individuals as either TRI-positive (P i.e. recent) or TRI-negative (N i.e. non-recent). Positive and negative in this context should not be confused with HIV-positive and HIV-negative.

The interaction between the virus and the immune system is complex, and individuals vary in their response to infection as assessed by a particular TRI. Modest variation is not intrinsically problematic, but serious complications arise if, in some individuals, the immune response is such that they remain indefinitely classified as TRI-positive or if individuals revert back to a TRI-positive classification as a result of advanced disease or in the presence of antiretroviral therapy. Unfortunately, both these complications arise for TRIs currently in use. This not only limits the applicability of the simple incidence estimator above, but also makes it difficult to define and estimate the mean duration spent in the recently infected state (i.e. to evaluate D). Methods for 'adjusting' estimates of incidence have been proposed [63, 40] and adopted by the United States Centers for Disease Control and Prevention [24] but are currently under debate [15, 39, 62, 116]. Recently, a formally rigorous framework

[1] This follows directly from the classical "Prevalence = Incidence × Duration" relationship.

has been developed [71, 69]. We provide a summary of the framework and explore its implications for the analysis of surveys and development of new TRIs.

5.2 Theoretical Framework

We now briefly describe the theoretical framework and how it can be generalised. The key results that emerge from the analysis are:

- A TRI is ideal if *all* individuals eventually progress permanently out of the TRI-positive state before there is any disease-related mortality. In this case, the TRI-positive category directly corresponds to a useful definition of 'recently infected' [51, 71], which means that an estimate for the number of recent infections is:
$$R = P.$$

- For a non-ideal TRI (i.e. when some individuals never progress out of the TRI-positive state), it is in principle still possible to estimate the number of individuals in a well-defined 'recently infected' state, even though this state is not directly observable in all individuals. If $\mathbb{P}_{\text{NP}}$ is the *proportion of the HIV-positive individuals who never progress on the TRI under consideration*, then an estimate for the number of recent infections is [71]:
$$R = P - \frac{\mathbb{P}_{\text{NP}}}{1 - \mathbb{P}_{\text{NP}}} N. \tag{5.1}$$
When the TRI is ideal, then $\mathbb{P}_{\text{NP}} = 0$, and this formula reduces to the previous expression.

- For all applications (including determination of a trend without regard to the absolute level of incidence), an estimate of $\mathbb{P}_{\text{NP}}$ is required.

- To determine the absolute level of incidence, it is also necessary to estimate the *mean time spent TRI-positive in the subset of individuals who eventually do progress to become TRI-negative*. This quantity, which we denote by ω, is analogous to the duration D in the simple estimator, but differs in the requirement that it should be estimated in the subset of individuals that progress on the TRI.

- As $\mathbb{P}_{\text{NP}}$ increases (i.e. a larger fraction of individuals fail to progress on the TRI) and as ω decreases (i.e. individuals spend less time in the TRI-positive state) statistical power is lost. This means that estimates of incidence will have more uncertainty (i.e. wider confidence intervals), and it is less likely that a true change in incidence will be detected.

Previous work by McDougal and colleagues [63], used terminology usually employed to characterise the performance of diagnostic tests, such as sensitivity and specificity, to characterise TRI performance. 'Recent infection' was defined as being infected for less than a particular time (chosen to be the mean window period). A sensitivity and two specificity parameters were introduced to characterise imperfect classification. No procedure incorporating the effect of parameter uncertainty has thus far been proposed to estimate statistical error or power for the McDougal approach. It has recently been shown that use of sensitivity and specificity parameters is a redundant description of the TRI characteristics [69, 117]. In contrast, the new framework defines the condition of being 'recently infected' directly in terms of the TRI result. This approach is applicable under less restrictive assumptions, is less prone to bias, and admits an equally informative description of TRI performance using only ω and $\mathbb{P}_{\text{NP}}$ [69].

In deriving the results outlined above, two assumptions were made. Firstly, it was assumed that individuals who do not progress on the TRI have the same survival outcomes as TRI progressors. There is, however, evidence for some TRIs that individuals that fail to progress on the test have a survival advantage[2]. Secondly, it was assumed that TRI progressors never regress back to the TRI-positive state, but there are indications that this is not true for some TRIs[3]. When these assumptions are true, $\mathbb{P}_{\text{NP}}$ is always equal to the proportion of non-recently infected individuals who are classified TRI-positive. When the assumptions are violated, this proportion, or false-recent rate, denoted by ε, varies according to the historic trajectory of the epidemic[4] [69, 36]. It is, however, still possible to estimate the number of recent infections by replacing $\mathbb{P}_{\text{NP}}$, in the expression (5.1) above, with an estimate of ε applicable to the time and place of an incidence survey [10] (See Appendix 5.6 for justification). The incidence estimator can then be written as:

$$I = \frac{P - \frac{\varepsilon}{1-\varepsilon}N}{\omega S}. \tag{5.2}$$

The inputs to this estimator are of two types: survey counts (P, N and S), which need to be estimated in every incidence survey, and parameters that describe the characteristics of the TRI (ω and ϵ), which ideally are estimated in a smaller number of parameter estimation studies.

[2] For example, in Baltimore, USA, 60% of elite suppressors (individuals with naturally suppressed virus below 50 copies per ml) failed to progress on the BED assay [57], and elite suppressors have been observed to survive for longer than others [45].

[3] For example, the rate of misclassification by the BED assay is observed to be higher in individuals with advanced infection [61] and individuals on ART [61, 41, 43].

[4] This would be consistent with the apparently higher BED assay false recent rate in Uganda [52] (an older, declining epidemic) than in South Africa [10] (a younger, growing epidemic) [36].

When ε and ω are known with sufficient accuracy, there are no *theoretical* reasons why an imperfect TRI should not allow the accurate estimation of incidence. However, two distinct types of *practical* problems arise—counting error and TRI parameter error. An important component of recent developments is the first consistent analysis of incidence uncertainty accounting for both counting and parameter error (see Appendix 5.6 for a description of the uncertainty expression). We now illustrate this uncertainty with a somewhat idealised model of the BED assay, which has received much attention and application [8].

5.3 Counting Error

Even in the largest HIV epidemics, infection events are relatively rare (about 2% of the population per year) and 'recent' infections (infections in the last 155 days or so, for the BED assay [63, 20]) are even less common (about 0.85% in a cross-section of the population). Thus, estimates of incidence are associated with substantial uncertainty since there are few recent infections to be counted. Figure 5.1 shows the coefficient of variation (CV)[5] for the estimator (5.2) calculated under various survey sample sizes and steady-state HIV incidence rates (See Appendix 5.6 for a description of the uncertainty and steady-state calculations). The TRI parameters (ω and ε) are assumed to be known with absolute certainty. Low values of CV are desirable and indicate that estimates of incidence have small confidence bounds, while high values indicate that incidence estimates will be less certain. For example, in a cross-sectional survey of 5000 individuals from a population with a steady-state incidence of 2.0 per 100 person years at risk (pyar) the CV is 25.8%, i.e. the 95% likelihood interval for an incidence estimate is 1.0–3.0 per 100 pyar.

To explore the ability to detect a change in incidence, a substantial reduction (halving) in incidence is simulated (initially in a steady-state epidemic, with prevalence remaining constant between the two surveys), and a two-tailed test of the null hypothesis that incidence is the same in the two surveys is performed. The possible outcomes are: sustaining the null hypothesis, or concluding that incidence has either increased or decreased. Figure 5.2 shows the probability of correctly inferring a reduction in incidence, when testing the null hypothesis at a significance level of $\alpha = 5\%$. A probability close to 100% indicates that reductions in incidence will be reliably detected, with a probability of less than 90% indicating that results will be unreliable. The South African National Strategic plan for HIV AIDS [95] has ambitiously set a target of halving incidence between 2007 and 2012. Our calculations suggest that the sample size of each of two surveys (in 2007 and 2012) required

[5] A coefficient of variation is the ratio of the standard deviation to the estimate.

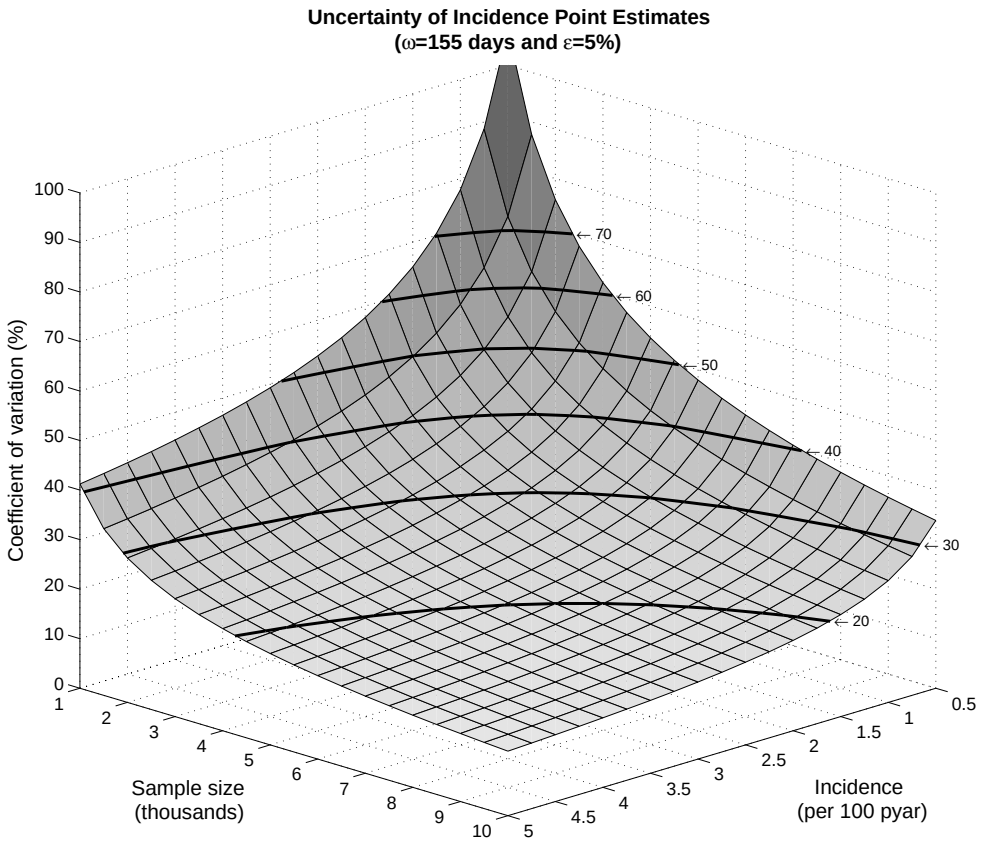


Fig. 5.1: The coefficient of variation of estimates of incidence using a TRI depends on the sample size of the survey and the true incidence rate. Note that a sample size of 10,000 approximates to the typical size of household-based surveys in sub-Saharan Africa, and that incidence in South Africa (where there is one of the largest epidemics) is estimated to be about 2 per 100 pyar. (Assumptions: $\omega = 155$ days; $\varepsilon = 0.05$; no TRI parameter uncertainty; steady-state epidemic conditions; mean survival with HIV: 11 years [31, 100].)

to reliably conclude that incidence has decreased, at the 5% significance level, is approximately 25,000.

5.4 TRI Parameter Error

In the previous section, it was assumed that the correct TRI parameters were known with certainty. The incidence estimates are very sensitive to changes in the values of ω and ε, however, and small differences between the values used in the calculation and the true values can lead to large errors. These parameters have to be estimated

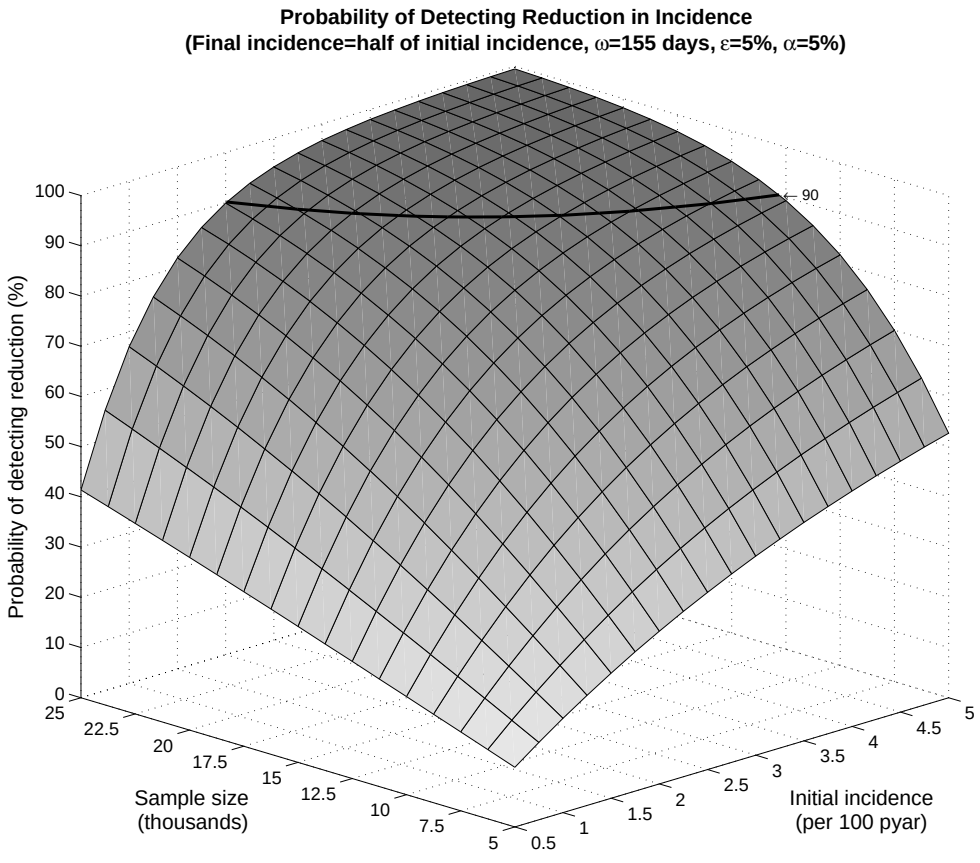


Fig. 5.2: The probability of detecting a reduction in incidence between two surveys, when incidence has actually been reduced by half, as a function of the sample size of the surveys (both assumed to be the same) and the baseline incidence rate. (Assumptions: $\omega = 155$ days; $\varepsilon = 0.05$;=0.05; no TRI parameter uncertainty; significance $\alpha = 5\%$; steady-state epidemic conditions at first survey, with equal prevalence at second survey; mean survival with HIV: 11 years [31, 100].)

in separate studies, usually using cohorts of individuals whose infection time is known approximately. Such cohorts are rare, however, and the numbers of individuals in them are typically small, resulting in substantial uncertainty for the values of ω and ε. In Figure 5.3 we explore the uncertainty of the estimator (expressed as a CV)[6], as a function of the uncertainty in the TRI parameters. For example, when the BED-like parameters are known with a CV of 15.0%, at a sample size of 5000 and a steady-state incidence of 2.0 per 100 pyar the CV, as a result of both counting error

[6] Reference to a CV for ω relates to the uncertainty of the estimate of ω, not the variation associated with progression times.

and parameter uncertainty, is 35.7%, i.e. the 95% likelihood interval for an incidence estimate is 0.6–3.4 per 100 pyar.

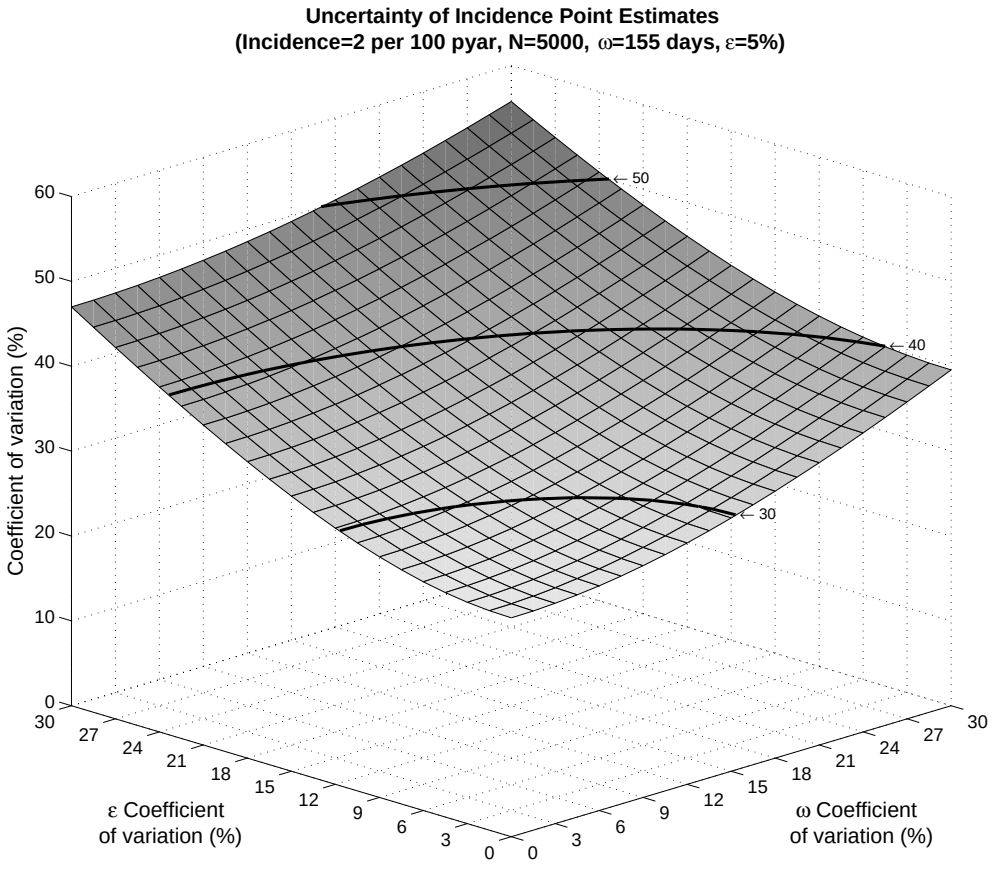


Fig. 5.3: Coefficient of variation of incidence estimator, using a BED-like assay on a sample size of 5,000, in a population exposed to an incidence of 2 per 100 pyar, as a function of the uncertainty in the TRI parameters, assumed to be normally distributed. (Assumptions: $\omega = 155$ days; $\varepsilon = 0.05$; steady-state epidemic conditions; mean survival with HIV: 11 years [31, 100].)

Since the TRI parameter estimation study may be conducted in a separate population, it is possible to introduce systematic bias if the true values of the TRI parameters vary between populations or over time. The few estimates of ε that have been published vary widely[7] presumably due to population differences in the historic courses of the epidemics, viral subtypes, host immune-profiles, and uptake of antiretroviral therapy. This undermines confidence in the ability to use an estimate for ε obtained in a different population to the one in which incidence is to

[7] For example, the false recent rate is estimated at 1.7% in a South African survey [10] and 26.7% in Rwanda and Zambia [52].

be estimated, and could contribute to the apparently inflated estimates of incidence reported recently [73, 75]. There is also currently no general theoretically unbiased procedure for estimating ε—work on this problem is in progress [65]. In Figure 5.4 we explore the systematic error in the incidence estimate, expressed as a percentage of the correct value, introduced by systematic errors in the TRI parameters, also expressed as percentages. There is a region in which bias may be small due to cancellation of systematic errors(see the zero error contour).

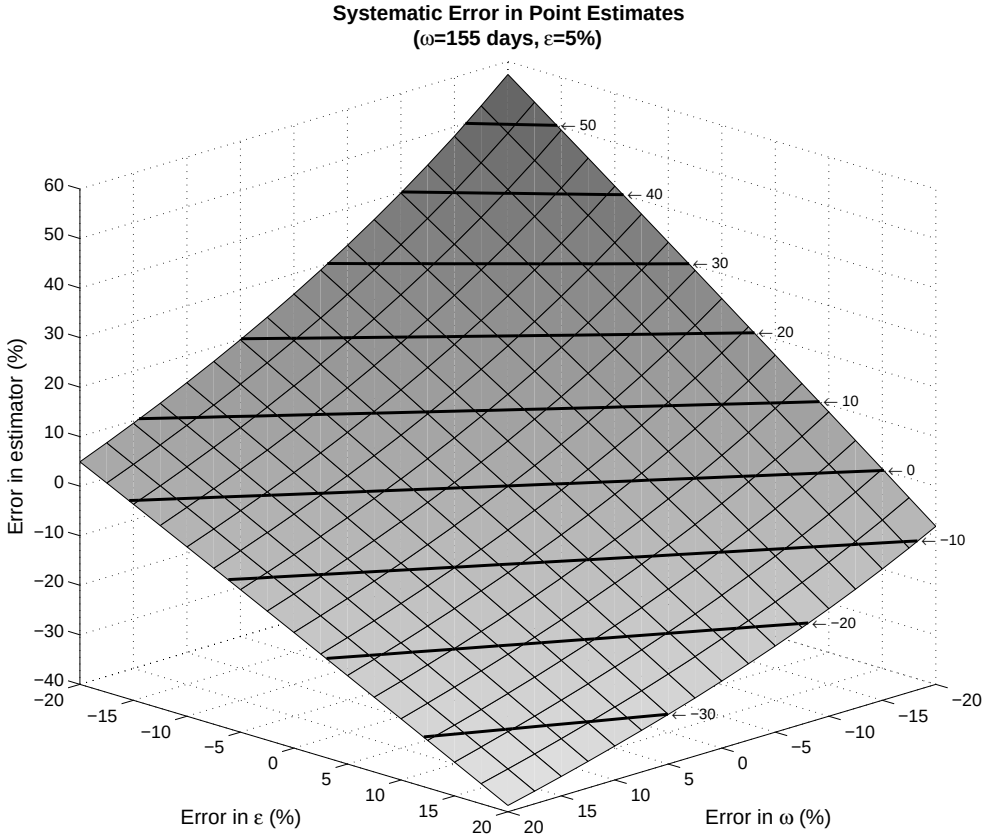


Fig. 5.4: Systematic error expressed as a percentage of the correct estimate, excluding counting error, observed in the incidence estimator, using a BED-like assay, as a function of a precisely known systematic error in the TRI parameters. (Assumptions: $\omega = 155$ days; $\varepsilon = 0.05$; steady-state epidemic conditions; mean survival with HIV: 11 years [31, 100].)

5.5 Conclusion

In the short-term, reports from early studies using BED should be interpreted with caution [8], given the substantial uncertainties identified above. Analysis of TRI data

should be performed within a more general theoretical framework [71, 69], rather than earlier methods. Most importantly, incidence surveillance should not currently rely on any single methodology, but make use of multiple methods for estimating incidence [30], such as interpretation of prevalence trends and epidemiological and demographic modelling [97, 37].

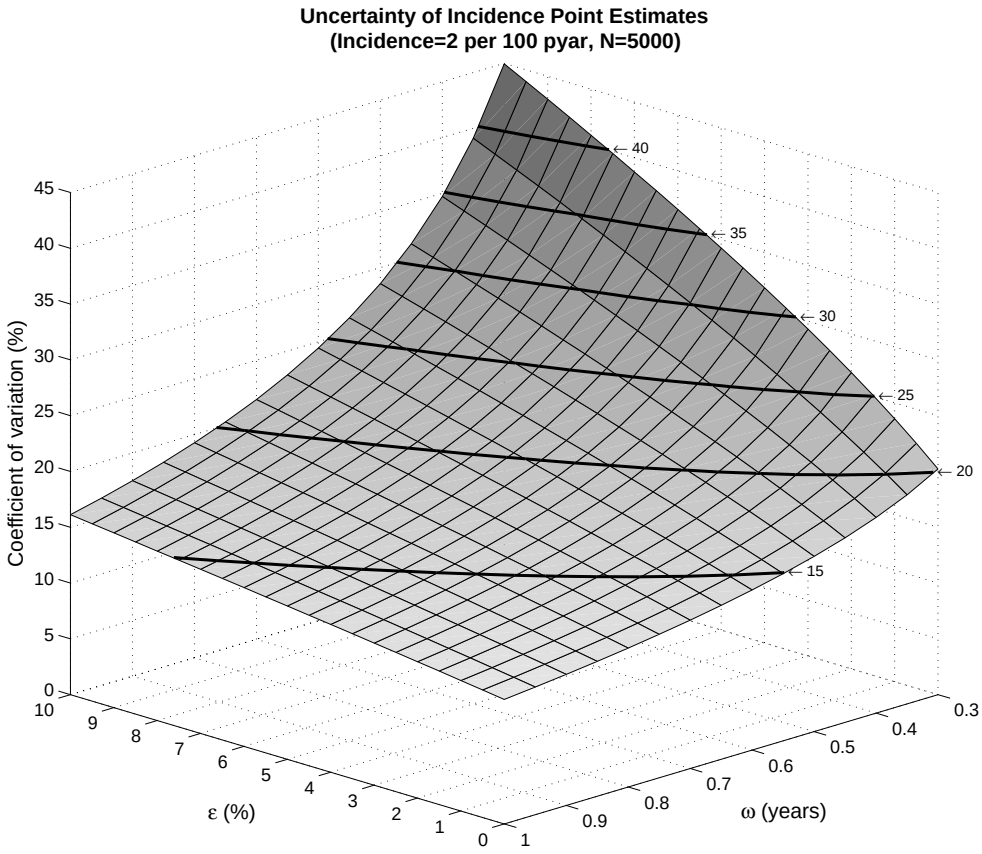


Fig. 5.5: Coefficient of variation of incidence estimator, on a sample size of 5,000, in a population exposed to an incidence of 2 per 100 pyar, as a function of the TRI parameters. (Assumptions: no TRI parameter uncertainty; steady-state epidemic conditions; mean survival with HIV: 11 years [31, 100].)

The search for robust means of estimating incidence from cross-sectional surveys is at a crucial juncture. Although an imperfect TRI can be used to estimate HIV incidence reliably, the reliance on having accurate and precise values of two key aspects of TRI performance (ω and ε) can undermine the use of this technology. The effect of ω and ε on statistical power is shown in Figure 5.5. While larger values of ω provide sufficient numbers of TRI-positive individuals to ensure statistical power, ω should not be so large that the estimated incidence is not representative

of the recent past. On this basis, a value of approximately six months to a year is desirable. It is also essential that ε be small[8]. Ideally, to ensure that the fraction of misclassifications is independent of time and epidemic state, inter-individual variability in TRI progression should be unrelated to survival outcomes, and there should be no regression to the TRI-positive state. These form the core requirements for the development of new TRI assays and algorithms used to estimate incidence.

In the next phase of TRI development, it will be essential to be guided by these insights into the key determinants of test performance, and to focus on characterising the performance of the test within a systematic framework.

5.6 Appendix

5.6.1 Justification for Replacing $\mathbb{P}_{\text{NP}}$ with ε

Denote the number of non-recently infected individuals that are incorrectly classified TRI-positive by F. The false recent rate ε is defined to be the fraction of all non-recently infected individuals that are classified as TRI-positive, which may be expressed in terms of F and the number of TRI-negative individuals, N, as:

$$\varepsilon = \frac{F}{N+F}.$$

This may be rearranged to provide an expression for the number of false recent results:

$$F = \frac{\varepsilon}{1-\varepsilon}N.$$

An estimate of the number of individuals that are *truly* recent, R, is the difference between the number of individuals that are TRI-positive, P, and the number of false recent results, i.e.

$$R = P - \frac{\varepsilon}{1-\varepsilon}N.$$

This is the same expression as before (5.1) with $\mathbb{P}_{\text{NP}}$ replaced by ε.

5.6.2 Uncertainty Expression

An expression for the uncertainty of the incidence estimator (5.2) may be derived using the Delta method [71]. Given the three survey counts (P, N and S), which are trinomially distributed, and the TRI parameters (ω and ε), which are assumed to be normally distributed (with coefficients of variation C_ω and C_ε), the expression for the coefficient of variation (CV) of the incidence estimator (C_I) is given by:

$$C_I = \sqrt{\frac{1}{N+P}\left(\frac{N+P+S}{S} + \frac{NP[1+\varepsilon/(1-\varepsilon)]^2}{[P-N\varepsilon/(1-\varepsilon)]^2}\right) + C_\omega^2 + \frac{C_\varepsilon^2\varepsilon^2N^2}{(1-\varepsilon)^4[P-N\varepsilon/(1-\varepsilon)]^2}}.$$

[8] Progress in this regard is being made, for instance using TRIs consisting of an assay in combination with clinical information [56].

This expression is used to compute 95% confidence intervals ($I \pm 1.96 \times C_I I$) and to generate the plots for Figures 5.1, 5.3 and 5.5 under the applicable assumptions. When no TRI parameter uncertainty is assumed (Figures 5.1 and 5.5) then C_ω and C_ε are set to zero. The hypothesis test simulation of Figure 5.2 consists of the following steps:

- The counts from two surveys are pooled to generate an estimate for the incidence implicit in the Null hypothesis;
- The above expression for C_I is used to generate a CV for the observed difference in the disaggregated incidence point estimates from the two surveys;
- The p value for the observed difference is computed and the Null hypothesis sustained/rejected according to a chosen level of significance.

The plot shows the probability of obtaining incidence differences consistent with inferring a reduction in incidence.

5.6.3 Steady-state Incidence

At equilibrium, the ratio of recent infections to non-recent infections is equal to the ratio of mean times spent in these categories. Under the assumption that there is no mortality in the recent category, this can be written as:

$$\frac{P - N\varepsilon/(1-\varepsilon)}{N + N\varepsilon/(1-\varepsilon)} = \frac{\omega}{\Omega - \omega},$$

where ω is the mean post-infection survival time [71].

The equilibrium prevalence is given by the product of the recruitment rate and the mean post-infection survival:

$$N + P = IS\Omega.$$

The above two equations together with the fact that the sum of P, N and S is equal to the total number of individuals recruited in the cross-section survey uniquely define the equilibrium counts.

Chapter 6

HIV Incidence in Rural South Africa: Comparison of Estimates from Longitudinal Surveillance and Cross-sectional cBED Assay Testing

∗ This chapter was coauthored with T. Bärnighausen, C. Wallrauch, A. Welte, N. Mbizana, J. Viljoen, N. Graham, F. Tanser, A. Puren and M.-L. Newell [10].

Abstract

Background: The BED IgG-Capture Enzyme Immunoassay (cBED assay), a test of recent HIV infection, has been used to estimate HIV incidence in cross-sectional HIV surveys. However, there has been concern that the assay overestimates HIV incidence to an unknown extent because it falsely classifies some individuals with non-recent HIV infections as recently infected. We used data from a longitudinal HIV surveillance in rural South Africa to measure the fraction of people with non-recent HIV infection who are falsely classified as recently HIV-infected by the cBED assay (the long-term false-positive ratio (FPR)) and compared cBED assay-based HIV incidence estimates to longitudinally measured HIV incidence.

Methodology/Principal Findings: We measured the long-term FPR in individuals with two positive HIV tests (in the HIV surveillance, 2003–2006) more than 306 days apart (sample size $n = 1,065$). We implemented four different formulae to calculate HIV incidence using cBED assay testing ($n = 11,755$) and obtained confidence intervals (CIs) by directly calculating the central 95th percentile of incidence values. We observed 4,869 individuals over 7,685 person-years for longitudinal HIV incidence estimation. The long-term FPR was 0.0169 (95% CI 0.0100–0.0266). Using this FPR, the cross-sectional cBED-based HIV incidence estimates (per 100 people per year) varied between 3.03 (95% CI 2.44–3.63) and 3.19 (95% CI 2.57–3.82), depending on the incidence formula. Using a long-term FPR of 0.0560 based on previous studies, HIV incidence estimates varied between 0.65 (95% CI 0.00–1.32) and 0.71 (95% CI 0.00–1.43). The longitudinally measured HIV incidence was 3.09 per 100 people per year (95% CI

2.69–3.52), after adjustment to the sex-age distribution of the sample used in cBED assay-based estimation.

Conclusions/Significance: In a rural community in South Africa with high HIV prevalence, the long-term FPR of the cBED assay is substantially lower than previous estimates. The cBED assay performs well in HIV incidence estimation if the locally measured long-term FPR is used, but significantly underestimates incidence when a FPR estimate based on previous studies in other settings is used.

6.1 Introduction

To understand the dynamics of the HIV epidemic and to target and evaluate interventions to prevent HIV infection, estimates of HIV incidence at the population level are of prime importance. HIV incidence estimates can be obtained through repeated HIV testing of individuals in longitudinal surveillances. Such surveillances, however, are difficult to establish and expensive to maintain. Longitudinal data on HIV status are thus rarely available [64]. Alternatively, HIV incidence can be estimated from changes in HIV prevalence over time. The validity of these estimates, however, depends on assumptions about survival time distributions among HIV-positive and -negative individuals, which are commonly quite uncertain [33, 120]. Finally, HIV incidence can be measured in a single cross-sectional survey using laboratory tests which distinguish recent from non-recent HIV infections, reducing the need for both longitudinal and repeated cross-sectional measurement in order to estimate HIV incidence [64].

In recent years, a number of large-scale cross-sectional HIV serosurveys have been conducted. For instance, between 2001 and 2008, 20 demographic health surveys (DHS) in developing countries have included nationally representative HIV serosurveys [27]. A valid and affordable laboratory procedure to distinguish between recent and non-recent infections would allow estimation of HIV incidence in these cross-sectional surveys. One serological method to differentiate recent from non-recent HIV infections uses the BED IgG-Capture Enzyme Immunoassay (cBED assay), which measures the proportion of HIV-1-specific IgG out of total IgG. This proportion should increase with time after HIV seroconversion [79]. Seropositive individuals who test below a certain threshold of this proportion (the BED threshold) are classified as recently infected, while those testing above the BED threshold are classified as non-recently infected [79]. The time period following seroconversion after which infections are no longer considered to be recent (the so-called window period of the cBED assay) is usually estimated at approximately half a year [79, 63, 40].

The cBED assay has been used to estimate HIV incidence in many countries,

including in Ethiopia [121], Rwanda [52], South Africa [86, 92], Uganda [73], Zambia [52], Zimbabwe [40], China [48, 58], and the United States [78, 35]. However, there has been concern that the cBED assay-based methods overestimate HIV incidence to an unknown extent because some non-recent infections are classified as recent [105]. In some individuals (so-called non-progressors) the proportion of HIV-1-specific IgG never rises above the recency threshold, and in other individuals (so-called regressors) who have been HIV-infected for a long time, the proportion may fall below the threshold after having previously progressed above it. Regression to levels below threshold can occur for a number of biological reasons that decrease HIV-1-specific IgG relative to total IgG, including viral suppression and immune reconstitution on antiretroviral treatment (ART), concurrent infections, and late-stage HIV disease [105]. It is in principle possible to account for non-recently HIV infected individuals who are misclassified as recently infected, but the HIV incidence estimates will depend on the estimate of a long-term false-positive ratio (FPR) [63, 40, 71]. All current methods for this correction effectively assume that by some finite time after HIV infection (the maximum BED progression time) all individuals, with the exception of non-progressors, will have progressed to the BED threshold [71]. From previous empirical observations, it is known that the maximum BED progression time is of the order of one year [63, 40]. Thus, the fraction of all people who have been HIV-infected at least as long as the maximum BED progression time who are below the BED threshold is the long-term FPR.

We use data from a large population-based longitudinal HIV surveillance to measure the long-term FPR in a rural African community with high HIV prevalence [119] and HIV incidence [9], and then compare HIV incidence estimates based on the cBED assay to estimates based on longitudinal HIV surveillance.

6.2 Methods

6.2.1 Setting

We used dried blood spot (DBS) specimens which were collected in the longitudinal population-based HIV surveillance conducted by the Africa Centre for Health and Population Studies (Africa Centre), University of KwaZulu-Natal [2]. The HIV surveillance area is located near the market town of Mtubatuba in the Umkhanyakude district of KwaZulu-Natal. The area is 438 square kilometers in size; it has a population of approximately 85,000 almost exclusively Zulu-speaking people who are members of about 11,000 households [99]. In 2004, the overall HIV prevalence among residents in the surveillance area was 27% in women (15 to 49 years of age) and 14% in men (15 to 54 years of age) [119]. The surveillance methods have

been described elsewhere [9, 7]. Ethics permission for the HIV surveillance at the Africa Centre was obtained from the Research Ethics Committee at the College of Health Sciences, University of KwaZulu-Natal. All participants in the study provided written informed consent for the analysis of their samples.

6.2.2 Samples

All women aged 15–49 years and all men aged 15–54 years who were resident in the surveillance area at the time of visit of an HIV surveillance fieldworker were eligible for HIV testing. Different samples were used for the different analyses conducted for this article. The samples for estimation of the long-term FPR consisted of cBED assay results for blood specimens contributed by individuals who tested HIV positive in the surveillance in the time period from June 2003 through June 2006. In order to be included in the sample, the specimens had to meet the following criteria. First, they were follow-up specimens from individuals who had previously tested HIV-positive in the surveillance. Second, the time period between the first positive HIV test and the follow-up specimen exceeded the maximum BED progression time. Third, the specimen was the earliest follow-up specimen that met the second criterion. Our count of long-term false-positive individuals included all individuals who were classified as recently HIV-infected and had been infected for longer than the maximum BED progression time, i.e. it included both non-progressors and regressors.

For the further cBED assay analyses we used a maximum BED progression time of 306 days (sample size $n = 1,065$) as baseline assumption. In order to assess the sensitivity of the long-term FPR to the assumed maximum BED progression time, we varied progression time length from 250 to 400 days in daily intervals. Table 6.1 shows sample size and the number of individuals who were falsely identified as recently HIV-infected for the BED progression times when the long-term FPR reaches its maximum and minimum and for all progression times in ten-day intervals from 250 to 400 days.

For the HIV incidence estimation based on longitudinal HIV status information, we included all individuals who tested at least twice for HIV in the period from June 2003 through June 2006 and whose first HIV test in this period was negative (4,869 individuals observed over 7,685 person-years). As in previous studies of HIV incidence based on data from longitudinal HIV surveillances [90, 32, 50, 22, 77], for the purpose of estimating exposure time, we used the mid-date between the last available negative HIV test and the first available positive HIV test as an estimate of the date of seroconversion. In addition, in order to test the robustness of the longitudinally measured HIV incidence estimates to changes in the assumption about serocon-

version dates, we re-estimated HIV incidence using the most extreme assumptions about the seroconversion date that are possible given the interval-censored information on seroconversion dates. At the one extreme, we assumed that all individuals in the longitudinal sample who seroconverted did so on the day immediately after the day of their last HIV-negative test. At the other extreme, we assumed that all individuals who seroconverted did so on the day of their first HIV-positive test. Under changes in the assumption of date of seroconversion, these two extremes yield maximum and minimum estimates of longitudinally measured incidence.

For the cross-sectional cBED-based HIV incidence estimation, we used the first available HIV test for all individuals tested in the time period January 2005 through June 2006 ($n = 11,755$), i.e. the period in which all second HIV tests of the people included in the longitudinal HIV incidence analysis took place. Thus, all 4,869 individuals in the longitudinal sample are also included in the sample for the cBED assay-based analysis.

6.2.3 Laboratory procedures

HIV status was determined by antibody testing with a broad-based HIV-1/HIV-2 enzyme-linked immunosorbent assay (ELISA; Vironostika, Organon Teknika, Boxtel, the Netherlands) followed by a confirmatory ELISA (GAC-ELISA; Abbott, Abbott Park, Illinois, USA) [7]. If HIV-positive status was confirmed, we used another spot from the same filter paper as used for the initial test in order to conduct the cBED assay (cEIA; CalypteH HIV-1 BED Incidence EIA, Calypte Biomedical Corporation, Maryland, USA). HIV-specific IgG were detected by the BED-biotin peptide, followed by a colour reaction with streptavidin-peroxidase. The optical density values were normalized in every run using a calibrator (normalized OD (ODn)=mean specimen OD/mean calibrator OD). Specimens with ODn less than or equal to 1.2 during an initial cBED screening test were confirmed by further cBED testing of the sample in triplicate. We took the median value of the three confirmatory test results as the final ODn value. As specified by the manufacturer, an HIV-1-positive specimen for which the cBED assay gave a final ODn of less than or equal to 0.8 was considered to be a specimen of recent HIV-1 infection. Otherwise, the specimen was classified as a non-recent infection [79].

6.2.4 Statistical analysis

Different formulae that use information obtained from the cBED assay have been proposed to estimate HIV incidence from cross-sectional surveys. These formulae provide incidence estimates expressed either as a rate, $\hat{I}_r$, (expressed, for instance, in

Maximum BED progression time	Sample size	No. of false-positive results	Long-term FPR (ε_2)	
(days)	(individuals)	(individuals)	Mean	95% CI
250	1100	18	0.0164	0.0097–0.0257
260	1094	18	0.0165	0.0098–0.0259
270	1090	18	0.0165	0.0098–0.0260
280	1083	18	0.0166	0.0099–0.0261
290	1081	18	0.0167	0.0099–0.0262
300	1070	18	0.0168	0.0100–0.0265
306	**1065**	**18**	**0.0169**	**0.0100–0.0266**
310	1056	18	0.0170	0.0101–0.0268
320	1043	18	0.0173	0.0103–0.0271
330	1035	18	0.0174	0.0103–0.0273
340	1017	18	0.0177	0.0105–0.0278
350	991	17	0.0172	0.0100–0.0273
360	**936**	**17**	**0.0182**	**0.0106–0.0289**
370	818	14	0.0171	0.0094–0.0285
374	**789**	**14**	**0.0177**	**0.0097–0.0296**
380	773	14	0.0181	0.0099–0.0302
390	755	14	0.0185	0.0102–0.0309
400	737	14	0.0190	0.0104–0.0317

Tab. 6.1: Long-term FPR. (FPR = false-positive ratio, CI = confidence interval. Row in bold font shows FPR at twice the window period of 153, 180, and 187 days, respectively.)

number of new HIV infections per 100 person-years) [71] or as the probability that in a given year a person will acquire HIV, i.e. an incidence proportion, $\hat{I}_p$, (expressed, for instance, in number of new HIV infections per 100 people per year) [63, 40]. Some of us have previously derived a formula from first principles to estimate HIV incidence based on the cBED assay [71], and have commented on the assumptions made in different formulae [69, 117]. Here, we implemented four different formulae found in the literature. The formula for HIV incidence derived by McDougal and colleagues (McDougal formula) [63] is

$$\hat{I}_p = \frac{fR}{fR + \omega N},$$

where R is the number of people who were classified as recently HIV-infected by the cBED assay and N is the number of individuals who tested HIV-negative. The mean window period of the cBED assay, ω, is the mean period of time from initial

seroconversion to reaching an ODn of 0.800 expressed in years in people who progress above the BED threshold [63]. The adjustment factor

$$f = [(R/P)\varepsilon_2]/[(R/P)(\sigma + \varepsilon_1 - 2\varepsilon_2)]$$

takes into account that the cBED assay does not have perfect specificity or sensitivity, P is the total number of people who tested HIV-positive, σ is the sensitivity of the cBED assay, ε_1 is the short-term FPR (i.e. over the period $[\omega, 2\omega]$), and ε_2 is the long-term FPR (i.e. over all times $> 2\omega$). Note that the short- and long-term specificities, ρ_1 and ρ_2, are related to the FPRs by $\rho_1 = 1 - \varepsilon_1$ and $\rho_2 = 1 - \varepsilon_2$, respectively. The formula of Hargrove and colleagues (Hargrove formula) [40] is

$$\hat{I}_p = \frac{R - \varepsilon_2 P}{R + \omega N - \varepsilon_2 (P + N)},$$

while the formula derived by McWalter and Welte (McWalter/Welte formula) [71] is

$$\hat{I}_r = \frac{R - \left(\frac{\varepsilon_2}{1-\varepsilon_2}\right)(P - R)}{\omega N}.$$

In addition, we implemented a simplified version of the McDougal formula. The adjustment factor used in the formula can be simplified to

$$f = [(R/P)\varepsilon_2]/[(R/P)(1 - \varepsilon_2)]$$

using the identity

$$\sigma + \varepsilon_1 - \varepsilon_2 = 1$$

which requires no more assumptions than are used by McDougal and colleagues [69, 117].

Note that in order to implement any of the above four formulae, estimates of the long-term FPR ε_2 and the window period ω are required. For our baseline estimation, we use an ω of 153 days, i.e. the window period that is recommended by the manufacturer of the commercially available cBED assay. Most previous studies reporting HIV incidence based on the cBED assay have used window periods between 150 and 160 days [63, 52, 73, 48, 58, 78, 35, 18, 34, 44, 59, 87]. A few studies have used a window period of 180 days [121, 86, 92], and a recent study from Zimbabwe calibrated a window period of 187 days in postpartum mothers enrolled in a Vitamin-A intervention trial [40]. In order to test whether our results are robust to changes in the window period estimate, we repeated our analyses with window periods of 180 and 187 days. The Hargrove and McDougal formulae require that the maximum BED progression time is twice the window period. The estimate of the long-term FPR thus depends on the choice of the window period (see Table 6.1).

Note also that the Hargrove, McWalter/Welte and simplified McDougal formulae do not require estimates of σ and ε_1, which—unlike ε_2—cannot be calibrated from longitudinal data if the intervals between the last negative and the first positive HIV test in seroconverters are of the order of one year [117]. The mean period of follow-up among seroconverters in our study was 1.4 years; we thus used estimates of σ (0.7680) and ε_1 (0.2770) from another study in order to implement the McDougal formula [63] (compare also [86]).

The McWalter/Welte formula expresses HIV incidence as a rate, i.e. as the number of HIV seroconversions per person-time at risk, while all other formulae express HIV incidence as an incidence proportion, i.e. the number of HIV seroconversions within a specified time period divided by the size of the population initially at risk. In order to directly compare all HIV incidence estimates in our study, we expressed the estimates based on the McWalter/Welte formula and the longitudinally measured HIV incidence both as rates (per 100 person-years) and as incidence proportions (per 100 people per year). We translated the rate estimates into proportions, assuming that the incidence rate, $\hat{I}_r$, is constant over time T, by using the relationship

$$\hat{I}_p = 1 - \exp(-\hat{I}_r T).$$

The authors of the four different formulae do not use equivalent methods for the calculation of confidence intervals (CIs). Thus, uncertainty analysis on the incidence estimates was performed as follows. Any observed proportion of HIV-negative, cBED-recent and cBED-non-recent individuals is an unbiased estimate of the underlying population proportions. Given an observed occurrence of the population proportions and the sample size, all attainable draws of the three counts can be enumerated and assigned their respective trinomial probability. Hence an exact cumulative probability distribution of attainable values of the incidence estimator can be computed. For each incidence estimate, we quote the estimator evaluated at the observed counts (the maximum likelihood estimate) and a confidence interval expressed as the central 95th percentile.

To control for differences in the sex-age composition between the sample used in the longitudinal HIV incidence estimation and the sample used in the cBED assay-based estimation, we weighted the sex- and five-year age group-specific longitudinal mean incidence rates by the proportions of individuals in each of the sex-age groups in the sample used for the cBED assay-based estimation

$$\hat{I}_{rs} = \sum_i w_{si} \hat{I}_{ri},$$

where $\hat{I}_{rs}$ is the sex-age adjusted mean incidence rate, w_{si} are the proportions of individuals in each sex-age group in the cBED assay sample, and $\hat{I}_{ri}$ are the sex-age

specific mean incidence rates. We estimated the variance of $\hat{I}_{rs}$, var($\hat{I}_{rs}$), as

$$\text{var}(\hat{I}_{rs}) = \sum_i w_{si}^2 \frac{\hat{I}_{ri}^2}{\hat{C}_i}$$

assuming that the number of HIV incident cases, $\hat{C}_i$, is Poisson distributed [29]. We calculated the 95% confidence limits for $\hat{I}_{rs}$ using the method based on gamma distributions described in Anderson and Rosenberg [4].

6.3 Results

Long-term FPR Counting the number of DBS specimens classified as recently HIV-infected by the cBED assay in the sample of all individuals who had a previous positive HIV test more than 306 days before the date of the cBED assay-tested specimen, we obtained a long-term FPR of 0.0169 (95% CI 0.0100–0.0266). When we varied the length of the maximum BED progression time from 250 to 400 days (in daily intervals), we found that the estimate of the long-term FPR did not change significantly over the time interval, with minimum and maximum long-term FPRs of 0.0164 (95% CI 0.0097–0.0257) and 0.0190 (95% CI 0.0104–0.0317), respectively (Table 6.1).

6.3.1 Incidence comparison

Of the 4,869 individuals included in the sample for longitudinal HIV incidence measurement, 224 people seroconverted in 7,685 person-years. Assuming that seroconversion occurred at the mid-date between the last available negative HIV test and the first available positive HIV test, longitudinally measured crude HIV incidence was 2.87 per 100 people per year (95% CI 2.53–3.27) (Table 6.2). Longitudinally measured HIV incidence increased to 3.09 per 100 people per year (95% CI 2.69–3.52), when we adjusted it to the age-sex distribution of the sample for the cBED assay-based incidence estimate.

Of the 11,755 individuals included in the sample for the cBED assay-based HIV incidence measurement, 9,236 tested HIV-negative and 2,519 HIV-positive. Of the individuals who tested HIV-positive, 165 were classified in cBED assay testing as recently HIV-infected and the remainder as non-recently infected. For given ε_2 and ω, the four different formulae to calculate HIV incidence from cBED assay measurement produced very similar results. Using the baseline estimate for ω of 153 days and the locally measured ε_2 of 0.0169, HIV incidence point estimates (per 100 people per year) varied between 3.03 (95% CI 2.44–3.63; McDougal formula) and 3.19 (95% CI 2.57–3.82; Hargrove formula) (Table 6.2). The cBED assay-based

HIV incidence estimates were thus very similar in magnitude and did not differ significantly from the estimates based on longitudinal measurement (crude and sex-age adjusted) (Table 6.2). Furthermore, when we implemented the cBED assay formulae using the lower bound or upper bound of the 95% CI of the locally measured long-term FPR (0.0100–0.0266), the cBED assay-based HIV incidence estimates did not differ significantly from the estimates based on longitudinal measurement. By contrast, when we implemented the cBED assay formulae using the externally measured long-term FPR of 0.0560 [63], all four cBED assay-based HIV incidence estimates were significantly lower than the longitudinal estimates (Table 6.2).

Our finding that the cBED assay-based HIV incidence estimate was not significantly different from the longitudinal HIV incidence estimate did not change when we applied the window periods of 180 and 187 days (and their corresponding long-term FPRs of 0.0182 and 0.0177 (see Table 6.1)). Using the McWalter/Welte formula, the cBED assay-based HIV incidence was estimated at 2.63 per 100 people per year (95% CI 2.10–3.18) with a 180-day window period and at 2.56 per 100 people per year (95% CI 2.04–3.08) with a 187-day window period. Neither of these estimates was significantly different from the longitudinally measured HIV incidence estimates or from the cBED assay-based incidence estimates based on a 153-day window period (see Table 6.2).

As described above, we conducted sensitivity analysis of the longitudinally measured HIV incidence estimate by changing the assumption about seroconversion dates. Assuming that all seroconverters became HIV-seropositive on the day following the last negative HIV test, crude HIV incidence was estimated at 2.97 per 100 person-years (95% CI 2.61–3.39). Assuming, on the other hand, that all seroconverters became HIV-seropositive on the day of their first positive HIV test, crude HIV incidence was estimated at 2.85 per 100 person-years (95% CI 2.51–3.25). The longitudinal HIV incidence estimates were thus highly robust to changes in the approach to computing the seroconversion date. Even under the most extreme possible assumptions, the mean HIV incidence changed by only 2% of the estimate based on the mid-date assumption, as reported in Table 6.2.

When we stratified HIV incidence by sex and five-year age group (starting at 15 years of age), we found that none of the cBED assay-based sex and age-specific estimates differed significantly from the corresponding longitudinally measured sex and age-specific estimates. However, our samples in each of the sex-age groups were too small to detect significant differences with reasonable confidence. The coefficients of variation (CVs) of the sex-age specific cBED assay-based HIV incidence estimates ranged from 18% to 203%; in 13 of the 15 sex-age groups the CVs were larger than 25%; in 10 sex-age groups the CVs were larger than 50%; and in 4 sex-age groups

Estimation type	Units	HIV incidence	
		Mean	95% CI
Longitudinal measurement (7,685 person-years, 224 seroconversions)			
Crude	(per 100 person-years)	2.91	2.56–3.32
Sex-age adjusted	(per 100 person-years)	3.14	2.73–3.58
Crude	(per 100 people per year)	2.87	2.53–3.27
Sex-age adjusted	(per 100 people per year)	3.09	2.69–3.52
cBED assay measurement (n = 11,755)			
Mean of locally measured long-term FPR ($\varepsilon_2 = 0.0169$)			
McWalter/Welte	(per 100 person-years)	3.22	2.57–3.87
McWalter/Welte	(per 100 people per year)	3.17	2.54–3.80
McDougal	(per 100 people per year)	3.03	2.44–3.63
Hargrove	(per 100 people per year)	3.19	2.57–3.82
McDougal, simplified	(per 100 people per year)	3.12	2.51–3.73
Lower bound of 95% CI of locally measured long-term FPR ($\varepsilon_2 = 0.0100$)			
McWalter/Welte	(100 person-years)	3.65	3.00–4.32
McWalter/Welte	(per 100 people per year)	3.58	2.95–4.22
McDougal	(per 100 people per year)	3.40	2.82–4.00
Hargrove	(per 100 people per year)	3.57	2.95–4.19
McDougal, simplified	(per 100 people per year)	3.52	2.91–4.14
Upper bound of 95% CI of locally measured long-term FPR ($\varepsilon_2 = 0.0266$)			
McWalter/Welte	(100 person-years)	2.60	1.96–3.27
McWalter/Welte	(per 100 people per year)	2.57	1.94–3.22
McDougal	(per 100 people per year)	2.49	1.89–3.11
Hargrove	(per 100 people per year)	2.63	1.99–3.29
McDougal, simplified	(per 100 people per year)	2.53	1.92–3.17
Externally measured long-term FPR ($\varepsilon_2 = 0.0560$)			
McWalter/Welte	(100 person-years)	0.65	0.00–1.33
McWalter/Welte	(per 100 people per year)	0.65	0.00–1.32
McDougal	(per 100 people per year)	0.66	0.00–1.33
Hargrove	(per 100 people per year)	0.71	0.00–1.43
McDougal, simplified	(per 100 people per year)	0.65	0.00–1.32

Tab. 6.2: HIV incidence estimates. (CI = confidence interval, FPR = false-positive ratio.)

they were larger than 100%.

6.4 Discussion

In a rural community in South Africa, we found a long-term FPR of the cBED assay of 0.0169. This value is substantially lower than the two previous estimates of the ratio. The first estimate (0.0560) was based on analysis of specimens from longer-term-infected individuals not known to have clinical AIDS, opportunistic infections, or to be on treatment in the USA [63]. The article, in which this value was published, provides neither the sample size for the measurement nor the confidence limits around the estimate [63]. Thus we cannot test whether the estimate is significantly different from the value that we measure in rural South Africa. The second estimate (0.0520) was based on specimens from 2,749 postpartum mothers enrolled in a Vitamin-A intervention trial in Zimbabwe [40]. This second estimate was significantly higher than the value measured in our study ($p < 0.0001$).

Many previous studies have used the first estimate of the long-term FPR in their estimations of HIV incidence based on cross-sectional cBED assay surveys (e.g. [52, 86, 73, 58]). In comparing cBED-based HIV incidence estimates to HIV incidence measured longitudinally in the same population, we have demonstrated that, had we used the long-term FPR of 0.0560, we would have significantly underestimated HIV incidence in this community. By contrast, using the locally measured ratio of 0.0169, we estimated an HIV incidence that does not differ significantly from the longitudinally measured incidence.

Our findings thus confirm the previous results by McDougal et al. [63] and Hargrove et al. [40] that cBED assay-based HIV incidence estimates are not significantly different from longitudinally measured HIV incidence, when a locally calibrated long-term FPR ratio is used to adjust for the imperfect long-term specificity of the cBED assay. At the same time, we have shown for the first time that the long-term FPR differs significantly across settings. Hence, results from studies that use a long-term FPR measured in another setting should be viewed with skepticism.

We further found that the different formulae to estimate HIV incidence based on the cBED assay results, did not produce significantly different values even though they differ in their underlying assumptions, suggesting that the choice of formula may not be very important for most practical purposes. Finally, we showed that the estimates of the long-term FPR based on data from a longitudinal HIV surveillance are very robust to changes in the definition of long-term (i.e. the choice of the maximum BED progression time).

Our longitudinal HIV incidence estimates in this article are slightly lower than previously published estimates from the same community [9], because the current study uses a sample that is different from the one used previously. In particular, unlike in the previous study, we excluded from the sample people who were identified

as members of a household in the study area, but who did not themselves live in the area. We excluded this population group (which faces a significantly higher risk of HIV acquisition than household members who live in the study area [7]), because cross-sectional cBED assay surveys usually do not trace such nonresident household members.

HIV incidence estimates by sex and age group are important for validating the cBED assay method as an approach to measure HIV incidence [40], and are an important disaggregation for health policy and planning, e.g. in order to inform the targeting of HIV prevention interventions. Our current sample lacked the statistical power to meaningfully stratify the HIV incidence estimates. As more data becomes available from our site, we will in the future analyze HIV incidence across population subgroups.

The promise of the cBED assay for HIV surveillance, program evaluation and policy making, lies in the fact that it allows HIV incidence estimation from cross-sectional samples. Cross-sectional HIV status information, however, does not permit estimation of the long-term FPR, requiring researchers to obtain this parameter independently. It is thus important that the parameters necessary for HIV incidence estimation are calibrated using data from those settings where longitudinal follow-up is available. A meta-analysis of the long-term FPR of the cBED assay may help explain why the parameter estimates differ and allow the determination of valid regional parameter estimates.

It may further be necessary to measure the long-term FPR repeatedly over time. For instance, one of the reasons why people with non-recent HIV infections are falsely classified as recently infected by the cBED assay is viral suppression due to ART [25]. In October 2004, ART started to become available through the public health services in the community in which this study took place. However, only a very small number of patients received ART during the study period. By the end of December 2005, i.e. half a year before the end of the study period, approximately 500 patients received ART through the public ART programme in the district in which this study took place. Because the HIV surveillance covers less than half of the district population, we estimate that in December 2005 less than 250 people in the surveillance area were receiving ART out of a total resident population of approximately 65,000 [42]. Future studies will need to investigate whether our locally estimated cBED long-term FPR changes with increasing ART coverage.

An alternative to using the long-term FPR in order to adjust cBED assay-based HIV incidence estimates for the presence of people who are falsely classified as recently HIV-infected is to use additional information on time since seroconversion to identify these individuals and correct the misclassification. Information on time

since seroconversion, which can be obtained in cross-sectional surveys, could be based on biological parameters that change with time since infection (such as CD4 count, total lymphocyte count, or viral load), clinical assessment (such as screening for HIV-related diseases that indicate late-stage HIV disease [123]), and screening for ART (through a question or laboratory test).

In conclusion, our study demonstrates that without a locally measured long-term FPR HIV incidence estimates based on the cBED assay may be severely biased, but that the cBED assay performs well in HIV incidence estimation, if a locally appropriate long-term FPR is used.

Chapter 7

HIV Incidence Estimation Using the BED Capture Enzyme Immunoassay: Systematic Review and Sensitivity Analysis

∗ This chapter was coauthored with T. Bärnighausen, Z. Rosner, M.-L. Newell and A. Welte [8], and is reproduced with permission from Wolters Kluwer Health (Rightslink licence no. 2597010845421): Epidemiology (2010) 21:685-97 DOI: 10.1097/EDE.0b013e3181e9e978.

Abstract

Background: HIV incidence estimates are essential for understanding the evolution of the HIV epidemic and the impact of interventions. Tests for recent HIV infection allow incidence estimation based on a single cross-sectional survey. The BED IgG-Capture Enzyme Immunoassay (BED assay) is a commercially available and widely used test for recent HIV infection.

Methods: In a systematic literature search for BED assay studies, we identified 1,181 unique studies, 1,138 of which were excluded based on titles or abstracts. We conducted reviews of the 43 remaining publications and a further 23 studies identified on conference web sites or by colleagues. Thirty-nine articles were included in the final review. We investigated the sensitivity of incidence values to various estimation methods and parameter choices.

Results: BED assay surveys have been conducted on five continents in general populations and high-risk groups, using one or more of ten distinct incidence formulae. Most studies used estimators that do not account for assay imperfection. Those studies that correct for assay imperfection commonly do not use locally-valid assay parameters. Incidence estimates were very sensitive to methodological and parameter choices. Most confidence intervals provided good assessment of uncertainty due to counting error, but only a few incorporated parameter uncertainty.

Conclusions: BED assay surveys can produce valid HIV incidence estimates, but many studies have not sufficiently accounted for assay imperfection. Future studies should (1) report all information necessary for incidence point and

uncertainty estimation, (2) use an unbiased estimator with locally-valid assay calibration parameters, and (3) compute confidence intervals that take into account parameter uncertainty.

7.1 Introduction

Estimates of HIV incidence—the rate of new infections in a population—are essential for monitoring the progress of HIV epidemics and for targeting and evaluating interventions that prevent HIV acquisition and transmission. Incidence estimates can be obtained through repeated testing of individuals in longitudinal surveillance. Such surveillances are, however, difficult to establish and costly to maintain; they may suffer from bias due to loss to follow-up [13] and they lack generalizability because participant behavior may change following risk-reduction counselling. An operationally-less-demanding approach to the estimation of incidence relies on tests that distinguish recent from non-recent infection in cross-sectional data.

While several tests for recent HIV infection have been developed [81, 64, 74, 98, 6], the BED IgG-Capture Enzyme Immunoassay (BED assay) has been frequently applied, especially in developing countries [92, 73, 48, 44]. From the time of the development of the BED assay in 2002 [79], there has been debate over how to correctly analyze the data generated by use of this assay. Recently, a number of authors have examined biases in the application of the BED assay using population models of HIV infection [15, 36, 116, 69]. Others have estimated the bias in large population-based surveys [11, 10].

We undertook a systematic review of the literature to survey the current practice in BED-assay application and to identify how the concerns regarding the accuracy of the test are being addressed. We present an overview of the current literature and collect information on methodological choices that are made in applying the assay. We show how sensitive the BED assay-based incidence estimates are to changes in methodology, including the incidence formula and calibration parameter values.

7.1.1 Basic Description of the BED Assay

The BED assay was developed by researchers at the US Centers for Disease Control and Prevention (CDC) for the purpose of identifying recently acquired HIV-1 infections regardless of viral subtype [79]. This was accomplished by producing a class-specific IgG antibody capture enzyme immunoassay (EIA) based on a trimeric branched peptide that includes gp41 immunodominant sequences from HIV-1 subtypes B, E and D—hence the name. The BED assay reports the proportion of HIV-1-specific immunoglobulin G (IgG) in total IgG as an optical density (OD) from

spectrophotometer measurements. To minimize the variations that occur in different runs, a normalized OD (OD-n) is determined using a calibrator specimen [79]. The proportion of HIV-1-specific IgG (and thus OD-n) increases with time after HIV infection.

Two major health organizations, the CDC and the United Nations Program on AIDS (UNAIDS), have issued statements regarding the use of the capture BED assay. While endorsing the assay for use in the US, the CDC recommendation [25] lists several situations in which the assay can produce false-recent results (i.e. non-recently infected individuals that are falsely classified as recently infected), including advanced HIV disease, chronic co-infection, and antiretroviral therapy. The statement concludes that "the BED HIV-1 Capture EIA was developed for and is solely used in the US in the context of HIV surveillance" and that the assay "may be less successful in a specimen-based system where [...] critical data cannot be ascertained." The most recent UNAIDS recommendation [105] concluded that "the BED-assay captures not only recent infections, but also late stage HIV infection (with or without antiretroviral therapy)" and that "[t]here is evidence that assay characteristics vary by HIV-1 subtype". UNAIDS thus called for "more research on the validity of the BED assay for estimating incidence" [105]. Neither the CDC nor UNAIDS have commented on the use of the assay since 2007.

7.1.2 Development of Incidence Estimators

To discriminate recent from non-recent infections, an OD-n threshold value (or cut-off) is chosen below which a specimen is classified as recently infected-specimens with an optical density above this value are classified non-recent. To estimate HIV incidence it is necessary to determine the mean length of time individuals remain classified as "recently infected" by the assay. This duration is usually called the "mean window period"; we denote it by ω. Estimating ω requires a calibration cohort study with frequent follow-up of individuals whose date of infection is approximately known. Using the well-known relationship between prevalence, incidence and duration, incidence is then estimated as the ratio of the sample count of recently infected individuals to the product of the count of susceptible individuals and the mean window period.

Following initial applications, it was discovered that the BED assay is an imperfect test, misclassifying some proportion of non-recently-infected individuals as recent. Two strategies have been used to correct for this shortcoming. The first corrects for assay imperfection on the level of the individual by using additional information (e.g. antiretroviral therapy [ART] utilization, AIDS diagnosis and previous HIV testing) to either re-classify or exclude individuals who are classified as

recent by the BED assay but are obviously non-recently infected (false-recent individuals). The second strategy corrects for assay imperfection at the population level, using incidence estimators that account for imperfect specificity of the BED assay.

A number of estimators have been proposed. In order to structure our review, we categorize the estimators into three "generations" (Figure 7.1). First-generation approaches include all simple applications of the prevalence-incidence relationship described above, making no attempt to account for false-recent individuals [79, 17, 47]. This approach results in overestimates of incidence if a substantial number of people are incorrectly classified as recent. The second-generation approach was initiated by McDougal and colleagues [63]. To account for BED assay imperfection, they introduced three additional assay-calibration parameters, i.e. sensitivity (σ), short-term specificity (ρ_1) and long-term specificity (ρ_2). The third-generation approach builds on the second-generation approach by simplifying the expressions. Whereas the second generation formulae depend on four parameters (ω, σ, ρ_1 and ρ_2), the third-generation approaches require only the window period and a false-recent rate (ε), which can be expressed in terms of the long-term specificity, $\varepsilon = 1-\rho_2$. Alternative names for the false-recent rate include false-positive rate [63], and false-positive ratio [10]. Note that while most authors refer to this parameter as a "rate", it is in fact the ratio of two counts (the number of persons with long-standing infection classified incorrectly as recent, and the total number of individuals with long-standing infection). Hargrove et al. [40] provided a new estimator that is equivalent to the McDougal estimator under the assumption that the sensitivity is equal to the short-term specificity. Later, Welte et al. [117] showed a formal mathematical relationship, different from Hargrove's assumption, between sensitivity, short-term specificity, and long-term specificity. This insight allows a consistent reduction of the McDougal estimator, and highlights the fact that the calibration parameters of the McDougal approach are not independent—ω and ε provide an equally-precise characterization of the performance of the assay when compared with the four parameters of the McDougal estimator. This over-parameterization in the McDougal approach may introduce unnecessary statistical uncertainty if the parameters are estimated independently, and makes it difficult to characterize the uncertainty in a consistent manner. McWalter and Welte [71] have derived a formally-consistent incidence relation that depends on fewer assumptions than either the McDougal or Hargrove approaches. More recently it has been shown that the approach of McWalter and Welte, when compared with the other third-generation approaches, is the only one that produces an unbiased estimate of incidence under the assumption of a steady state epidemic [69].

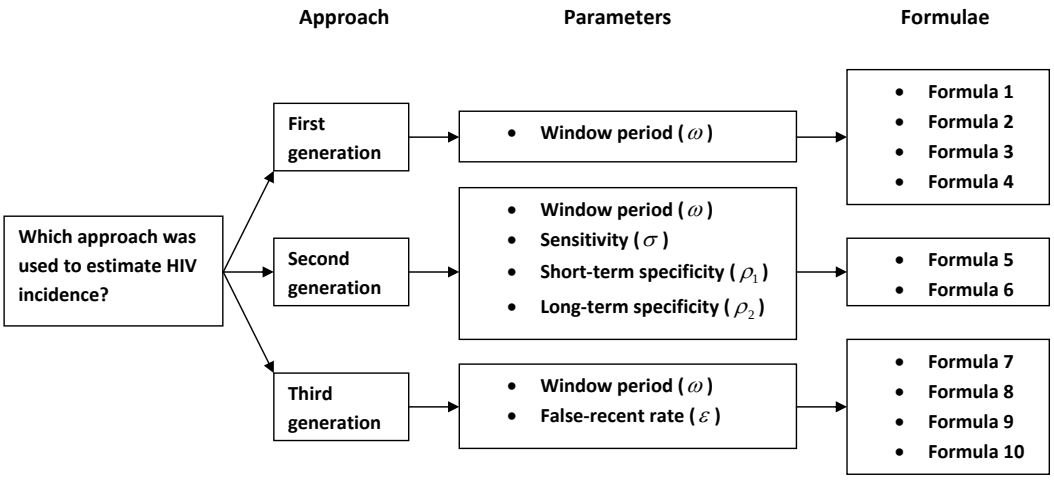


Fig. 7.1: Overview of approaches, required parameters and formulae used to estimate HIV incidence from cross-sectional surveys using the BED assay.

7.2 Methods

7.2.1 Literature Search Strategy

We carried out a systematic literature search in the PubMed electronic database [85]. To identify articles, we combined search themes using the Boolean operators "and" and "or": HIV "and" (BED assay "or" recent infection). Wherever possible, we drew search terms for each theme from the Medical Subject Headings (MeSH) [72], the controlled vocabulary used for subject indexing in PubMed:

```
("HIV"[MeSH] OR "HIV-1"[MeSH] OR "HIV-2"[MeSH] OR
"HIV Seroprevalence"[MeSH] OR "HIV Seropositivity"[MeSH])
AND
("IgG capture"[All Fields] OR "BED"[All Fields] OR
"CEIA"[All Fields] OR "EIA"[All Fields] OR
"IgG immunoassay"[All Fields] OR "immunoglobulin G"[MeSH] OR
"recency"[All Fields] OR "recent infection"[All Fields] OR
"incidence"[MeSH]).
```

We used all MeSH terms in their "exploded" versions so that all narrower terms categorized below each selected term in the vocabulary hierarchies were also included in the searches. In addition, we searched for terms that did not exist in MeSH using the "All Fields" category of the PubMed electronic database.

The development of the BED assay was first described by Parekh et al. [79] in a publication dated March of 2002. To ensure that we included all articles describing

studies using the assay, we searched for articles published on or after the 1st of March 2000, i.e. two years prior to the publication describing the development. Our search period ended on the 4th of March 2009. In addition to PubMed, we searched the websites of the Conference on Retroviruses and Opportunistic Infections (CROI) (covering all CROI from January 1997 to March 2009) [1] and the International AIDS Society (IAS) (covering all Conferences on HIV Pathogenesis and Treatment, and all International AIDS Conferences (AIDS) from 2001 through 2008) [46] for abstracts containing the terms "BED", "cBED", "CEIA", "EIA", "immunoglobulin G", "IgG immunoassay", "recency" and "recent infection". These terms were also used to search the National Library of Medicine Gateway (NLM Gateway) [76], which includes abstracts from twenty-nine HIV-related conferences [76]. We further searched the reference lists of reviews, editorials, commentaries and all publications included in the final review. Finally, we asked colleagues with a research interest in HIV epidemiology or prevention to identify studies that report findings based on BED assay surveys.

Our initial PubMed search identified a total of 1181 unique studies, 1138 of which were excluded based on titles or abstracts. Studies were excluded at this stage if they did not report HIV incidence estimates, reported only HIV incidence estimates that were not based on BED assay surveys, or did not specify the populations in which a BED assay-based HIV incidence estimates were obtained. Studies were further excluded if they were not written in English, or were reviews, letters, editorials or commentaries. We conducted full-text reviews of the forty-three remaining publications and twenty-three studies identified in conference abstract databases, in NLM Gateway, or by colleagues. The only reason for exclusion after full-text review was that studies reported only HIV incidence estimates that were not based on BED assay surveys. We did not identify any studies through screening of references that were not also identified in one of the other searches. Four conference abstracts were excluded because they reported data contained either in a full-text article or in an abstract with a later publication date.

We identified thirty-nine studies for the final review—a summary is provided in Figure 7.2.

7.2.2 Sensitivity Analysis

In order to explore the robustness of the incidence estimates to various methodological and parameter choices, we performed a number of sensitivity analyses. For each of the full-text articles that reported the formula, parameter values and survey counts used in incidence estimation, we recomputed incidence. To explore the sensitivity with respect to choices of methodology, we computed incidence using

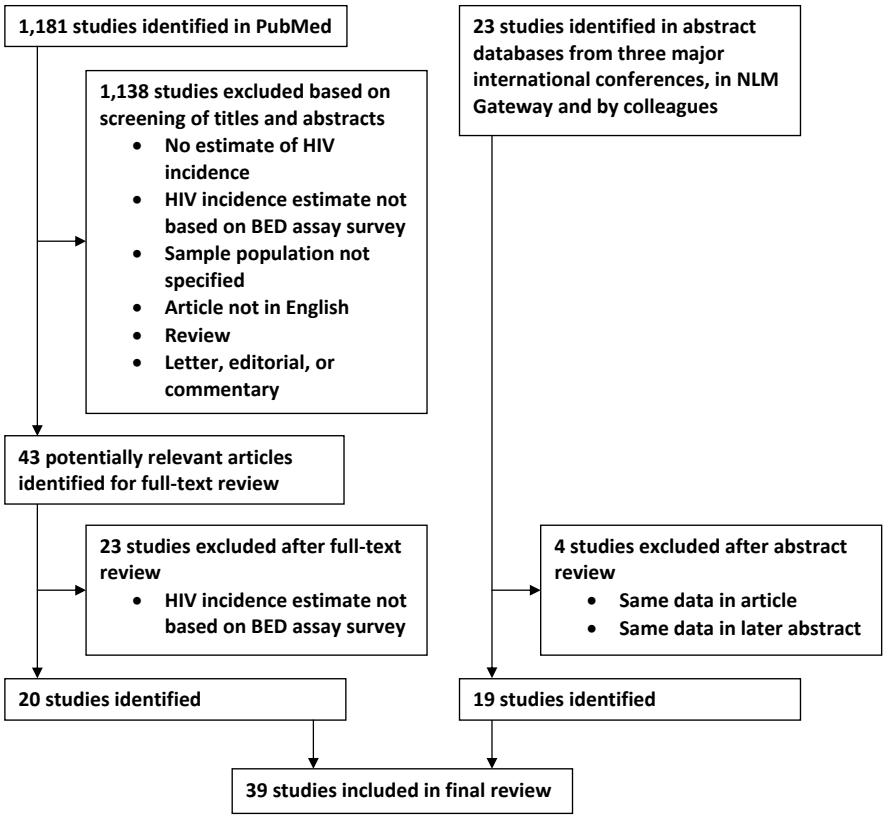


Fig. 7.2: Flow chart of search and selection criteria for studies included in the final review.

all the different estimators found in the literature, in each of the three generations mentioned above.

It has been noted that the incidence estimate is sensitive to the choice of calibration parameters [10, 40]. In particular, a locally-valid estimate of the long-term specificity is necessary because overestimates and underestimates of incidence can occur when false-recent results are not appropriately accounted for. For this reason, we explore the sensitivity of the incidence estimates to changes in both the window period and the false-recent rate (alternatively, the long-term specificity).

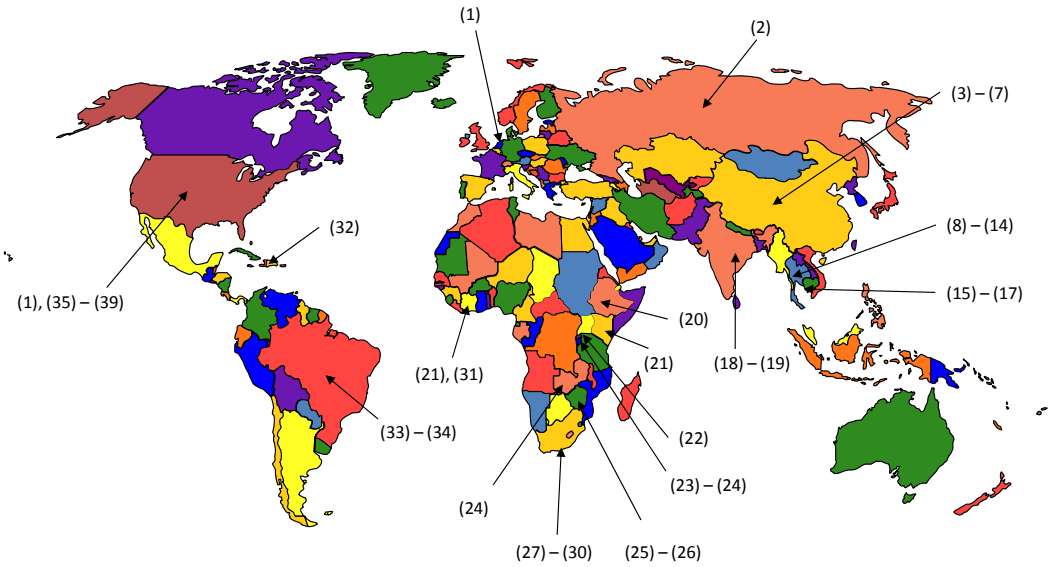
When presenting incidence estimates, it is important that the variability of the estimate be stated as a confidence interval (CI). We computed CIs under a variety of assumptions in order to explore the effect of the parameter uncertainty, using a closed-form expression for the coefficient of variation of the estimator derived with the delta method [71], which assumed parameter error was distributed normally.

For the calibration-parameter and CI sensitivity analyses, we used the incidence estimator of McWalter/Welte, because it is the least biased of the estimators [69]

and is the only estimator with an expression for standard deviation that incorporates the effect of parameter uncertainty [71].

7.3 Results

From each of the thirty-nine studies, we extracted information on study characteristics and results, as displayed in Table 7.1 and Figure 7.3 and described below. Additional survey data, only available for the full-text articles, is shown in Table 7.2 and described below.



(1) United States, The Netherlands (MSM, high-risk women) 1998-1999; **(2)** Russia (IDU) NR; **(3)** China (IDU) 2000-2006; **(4)** China (IDU) 2002-2005; **(5)** China (MSM) 2005-2006; **(6)** China (CSW) NR; **(7)** China (CSW) NR; **(8)** Thailand (military) 1991; **(9)** Thailand (IDU) 1996; **(10)** Thailand (IDU) 1999-2000; **(11)** Thailand (ANC, CSW) 2004-2005; **(12)** Thailand (military) 2005-2006; **(13)** Thailand (GP) NR; **(14)** Thailand (military, IDU) NR; **(15)** Cambodia (CSW, ANC, police) 1999-2002; **(16)** Cambodia (GP) 2006; **(17)** Cambodia (fishermen) NR; **(18)** India (STD) 2002-2004; **(19)** India (IDU) NR; **(20)** Ethiopia (ANC) 1995-2003; **(21)** Ivory Coast, Kenya (ANC, GP) 1998-2004; **(22)** Uganda (GP) 2004-2005; **(23)** Rwanda (ANC) 1989-1993; **(24)** Rwanda, Zambia (discordant couples) 2004; **(25)** Zimbabwe (postpartum mothers) 1997-2000; **(26)** Zimbabwe (GP) NR; **(27)** South Africa (GP) 2003-2006; **(28)** South Africa (GP) 2005; **(29)** South Africa (GP) 2005; **(30)** South Africa (GP) 2006-2007; **(31)** Ivory Coast (blood donors) 1997-2003; **(32)** Dominican Republic (CSW) 2004-2005; **(33)** Brazil (GP) 2004-2005; **(34)** Brazil (MSM) 2004-2005; **(35)** United States (ANC) 1991-1998; **(36)** United States (MSM) 2000-2003; **(37)** United States (MSM, STD, non-IDU) 2002-2004; **(38)** United States (STD) 2004-2005; **(39)** United States (GP) 2006.

Key: MSM indicates men who have sex with men; IDU, intravenous drug users; NR, not reported; CSW, commercial sex workers; ANC, antenatal care clinic attendees; GP, general population; STD, sexually transmitted disease clinic attendees.

Fig. 7.3: Map showing location, population and observation period of 39 studies.

Location, Population Setting, Study Type and Assay

Thirty-nine relevant studies were published between 2003 and 2009 (twenty full-text articles and nineteen abstracts) (Figure 7.2 and Table 7.1). In these studies, the BED assay was used to assess HIV incidence in regions throughout the world. Seventeen studies used data from Asia [44, 88, 89, 96, 103, 111, 110, 113, 114, 124, 48, 82, 104, 58, 49, 5, 94], twelve from Africa [19, 40, 52, 73, 86, 87, 92, 93, 121, 10, 101, 55], six

Authors	Pub date	Observation period	Country	Population	Study type	Incidence	Total sample size (range of sizes)	OD-n cut-off	Generation of formula	CI method
Apornpong et al.	2007†	NR	Thailand	GP	IS	5.8%	435#	0.8	NR	NR
Bärnighausen et al.	2008	2003-2006	South Africa	GP	LS	3.17%	11755	0.8	2nd, 3rd	Trinomial exact
Buchacz et al.	2008	2004-2005	USA	STD	IS	9.5%	457, ES-2, IE(M, T, X)	0.8	1st	Normal (R)
Bultreys et al.	2004†	1989-1993	Rwanda	ANC	CC	5.7%	5866	NR	NR	NR
Gupta et al.	2007	2004-2005	Dominican Republic	CSW	CC	1.0%	482, ES-3, IE(M)	0.8	1st	Poisson/Bonferroni
Hall et al.	2008‡	2006	USA	GP	RS	0.0228%	NR, IE(A, X, T)	0.8	NR	Delta method
Hargrove et al.	2008	1997-2000	Zimbabwe	Postpartum mothers	CC	6.0%	14110, ES-53	0.8	1st, 2nd, 3rd	Delta method
Hu et al.	2003	1996	Thailand	IDU	PS	17.3%	1969	1.0	1st	Poisson/Bonferroni
Jiang et al.	2007	2000-2006	China	IDU	SS	0.57-9.58%	17213 (251-3244), IE(A)	0.8	1st	Normal (R)
Kana et al.	2008†	2005-2006	Thailand	Military	IS	0.15-0.22%	87178 (27706-29858)	0.8	NR	NR
Karita et al.	2007	2004	Rwanda, Zambia	Discordant couples	PS	3.4-6.1%	6436 (1000-2004)§	0.8	1st, 2nd, 3rd	Normal (R)
Kim et al.	2007†	1998-2004	Ivory Coast, Kenya	ANC,GP	SS	2.2-3.1%	NR	NR	NR	NR
Li et al.	2008	2005-2006	China	MSM	IS	2.9-3.6%	1067 (526-541), IE(A, X)	0.8	2nd	NR
McDougal et al.	2006‡	1998-1999	USA, Netherlands	MSM, high-risk women	CC	2.91%	NR, IE(A, X, T)	0.8	2nd	Normal (R)
Mermin et al.	2008	2004-2005	Uganda	GP	CS	1.8%	18525§	0.8	2nd	Normal (R)
Nesheim et al.	2005	1991-1998	USA	ANC	IS	0.24%	48572 (3690-7624)§	1.0	1st	Modified Wald
Pilpat et al.	2006†	2004-2005	Thailand	ANC, FSW	SS	0.2-3.7%	62156 (2068-25308)	NR	NR	NR
Priddy et al.	2007	2002-2004	USA	MSM, STD, non-IDU	IS	1.3%	2202, IE(T)	0.8	1st	Normal (R)
Rehle et al.	2007‡	2005	South Africa	GP	CS	1.4%	15851	NR	2nd	Normal (R & N)
Sakarovitch et al.	2007‡	1997-2003	Ivory Coast	Blood donors	CC	6.2%	NR	1.0	1st	Normal (R)
Saphonn et al.	2005‡	1999-2002	Cambodia	CSW, ANC, police	SS	0.26-13.9%	39572 (62-4404)	1.0	1st	Poisson/Bonferroni
Saphonn et al.	2007†	2006	Cambodia	GP	IS	3.3%	1502	NR	NR	NR
Shisana et al.	2005‡	2005	South Africa	GP	CS	2.7%	15851	0.8	1st	NR
Simbayi et al.	2007‡	2006-2007	South Africa	GP	CS	1.5%	2565 (341-2224)	NR	2nd	Normal (R & N)
Sinthuwattaniwibool et al.	2005†	1991	Thailand	Military	CC	0.51-1.44%	1113	0.8	NR	NR
Srikrishnan et al.	2006†	NR	India	IDU	IS	4.66%	866	NR	NR	NR
Truong et al.	2004†	NR	Cambodia	Fishermen	IS	4.35%	400	0.75/1.0	NR	NR
Truong, Fritz et al.	2007†	NR	Zimbabwe	GP	CS	1.91%	1097	0.8	2nd	NR
Truong, Kellogg et al.	2007†	2000-2003	USA	MSM	IS	1.90-4.46%	15010 (NR-NR)	NR	2nd	NR
Truong et al.	2008†	2002-2004	India	STD	IS	6.81%	3403	NR	2nd	NR
Velasco de Castro et al.	2006†	2004-2005	Brazil	GP	IS	1.68%	9008	NR	NR	NR
Velasco de Castro et al.	2007†	2004-2005	Brazil	MSM	IS	9.12%	498	NR	NR	NR
Verevochkin et al.	2008†	NR	Russia	IDU	CC	12.7%	NR	NR	1st	NR
Wang et al.	2007†	NR	China	CSW	CS	2.1%	737	0.8	NR	NR
Wang & Xu	2008†	NR	China	CSW	IS	2.0-0.8%	737	NR	NR	NR
Wasinrapee et al.	2004†	1999-2000	Thailand	IDU	PS	7.4%	4943#	1.0	1st	NR
Wasinrapee et al.	2006†	NR	Thailand	Military, IDU	CC	1.1-7.5%	NR	NR	1st	Normal (R)
Wolday et al.	2006‡	1995-2003	Ethiopia	ANC	SS	1.0-6.0%	7744	NR	1st	Normal (R)
Xiao et al.	2007	2002-2005	China	IDU	IS	0.4-8.8%	16566 (1170-7307)	NR	1st	Log of Normal (R)

Key: OD-n indicates normalized optical density; CI, confidence interval; NR, not reported; GP, general population; STD, sexually transmitted disease; CSW, commercial sex workers; IDU, intravenous drug users; MSM, men who have sex with men; ANC, antenatal care clinic attendees; IS, stand-alone HIV incidence study; LS, longitudinal population-based HIV surveillance; RS, case-reporting HIV surveillance; SS, sentinel HIV surveillance; CC, clinical cohort study; CS, cross-sectional HIV survey; PS, preparatory study for clinical trial; ES-n, n excluded samples; IE(.), Inclusion/Exclusion criteria (M = HIV+ individuals with missing BED samples; A, AIDS diagnosis; X, antiretroviral treatment status; T, previous HIV test); R, BED+; N, HIV-.
†Abstract presented at a conference.
‡Full text article with insufficient data for incidence recalculation.
§Adjustment for HIV+ individuals with missing BED samples.
#Possible adjustment for HIV+ individuals with missing BED samples.

Tab. 7.1: Descriptive table.

Authors	Window period	Year length	Effective window	Sensitivity	Short-term specificity	Long-term specificity#	HIV- (N)	HIV+ (P)	BED+ (R)
Bärnighausen et al.	153	365.00	0.4192	-	-	0.9831	9236	2519	165
Buchacz et al.	155	365.25	0.4244	-	-	1	268	187	11
Gupta et al.	153	365.25	0.4189	-	-	1	463	16	2
Hall et al.†	156	365.00	0.4274	NR	NR	NR	NR	6864	2133
Hargrove et al.	187	365.00	0.5123	-	-	0.9480	9562	4495	517
Hu et al.	160	365.00	0.4384	-	-	1	1375	594	113
Jiang et al. (City D All)	155	365.00	0.4247	-	-	1	2811	433	25
Jiang et al. (City D IDU)	155	365.00	0.4247	-	-	1	585	275	25
Karita et al. (Masaka)	153	365.00	0.4192	0.7680	0.7230	0.9440	1191.8‡	151	39
Karita et al. (Kakira)	153	365.00	0.4192	0.7680	0.7230	0.9440	1752.6‡	190	47
Li et al. (Data for 2005)	155	365.00	0.4247	0.7682	0.7231	0.9443	509	17	7
Li et al. (Data for 2006)	155	365.00	0.4247	0.7682	0.7231	0.9443	515	26	9
McDougal et al.†	153	365.00	0.4192	0.7680	0.7230	0.9440	NR	NR	NR
Mermin et al.	155	365.00	0.4247	0.7680	0.7230	0.9440	16331.5‡	1023	172
Nesheim et al.	160	365.25	0.4381	-	-	1	48018	554	50.2§
Priddy et al.	153	365.00	0.4192	-	-	1	2136	66	12
Rehle et al.†	180	365.00	0.4932	0.7682	0.7231	0.9443	NR	NR	NR
Sakarovitch et al.†	160	365.00	0.4384	0.8570	0.7710	NR	NR	NR	NR
Saphonn et al.†	168	365.00	0.4603	-	-	1	NR	3599	NR
Shisana et al.†	180	365.00	0.4932	-	-	1	NR	NR	181
Simbayi et al.†	180	365.00	0.4932	0.7682	0.7231	0.9443	NR	NR	NR
Wolday et al.†	180	365.00	0.4932	-	-	1	6394	1350	NR
Xiao et al. (IDUs)	153	365.00	0.4192	-	-	1	945	225	34
Xiao et al. (County B)	153	365.00	0.4192	-	-	1	6482	825	116

Table does not include studies for which only abstracts were available because abstracts did not include the required information.

Key: NR indicates not reported.
†Insufficient data to recompute incidence.
‡Adjustment for HIV+ individuals with missing BED samples.
§Non-standard adjustment for HIV+ individuals with missing BED samples.
Long-term specificity = 1 − false recent rate. A specificity of 1 indicates use of a first-generation approach (i.e. a false-recent rate of 0).

Tab. 7.2: Calibration information and sample counts.

from North America [35, 78, 84, 63, 18, 102], two from South America [107, 108], and two from Europe [63, 109]. One study used data from more than one geographical region [63]. Eleven studies were conducted in the general population [89, 5, 73, 86, 92, 93, 10, 101, 55, 35, 108] seven in intravenous drug users [44, 96, 113, 114, 124, 48, 109], six in antenatal care (ANC) attendees [88, 82, 19, 121, 55, 78] five in commercial sex workers (CSW) or female sex workers [88, 111, 110, 82, 34], five in men who have sex with men (MSM) [58, 84, 63, 102, 107], three in the military [114, 49, 94], three in sexually transmitted disease (STD) clinic attendees [103, 84, 18], and one each in post-partum mothers [40], non-intravenous drug users [84], discordant couples [52], blood donors [87], fishermen [104], and police [88]. Six studies used data from two or more different populations [88, 114, 82, 55, 84, 63]; they are included in the counts of each population above. The countries in which the reviewed studies took place and the populations in which the BED assay surveys were conducted are shown in Figure 7.3. The number of published studies using the BED assay to estimate HIV incidence increased from one in 2003, to three in 2004, four in 2005, six in 2006, and fourteen in 2007, but decreased slightly to eleven in 2008 (Table 7.1).

The BED assay was applied to samples collected as part of case-reporting surveillance [35] (i.e. passive surveillance through which all individuals who are diagnosed as HIV-infected in voluntary counseling and testing centers are reported to a central organization, such as the US CDC), longitudinal population-based surveillance [10] (i.e. active surveillance in which eligible individuals contribute blood samples for an

HIV test repeatedly over time), sentinel surveillance [88, 48, 82, 121, 55], (i.e. active surveillance that collects blood samples for HIV tests from all individuals belonging to a certain population group, e.g. individuals attending one of a selected set of antenatal care clinics), clinical cohort studies [114, 94, 19, 40, 86, 63, 109, 34], preparatory studies for clinical trials [44, 113, 52], cross-sectional HIV surveys [111, 73, 86, 92, 93, 101], and stand-alone HIV incidence studies [89, 96, 103, 111, 124, 104, 58, 49, 5, 78, 84, 18, 102, 107, 108].

Total sample sizes in the reviewed studies ranged from 400 to 87,178 across the 33 studies that reported the sample size (Table 7.1). Nineteen studies reported using the commercially available BED immunoassay produced by Calypte Biomedical Corporation [21], while the rest did not report the manufacturer of the assay used.

Incidence Estimation Approach

Fourteen studies [44, 88, 113, 114, 124, 48, 87, 92, 121, 78, 84, 18, 109, 34] used only a first-generation approach, eight [103, 58, 73, 86, 93, 101, 63, 102] used only a second-generation approach; one [10] used second- and third-generation approaches; and two [40, 52] used formulae from all three generations. Fourteen studies [89, 96, 103, 111, 110, 82, 49, 5, 94, 19, 55, 35, 107, 108] did not report the formula used to estimate HIV incidence.

Three studies collected additional clinical information on study participants but did not consider using it to exclude or reclassify individuals [124, 52, 34]. Three other studies indicated that their samples were unlikely to include individuals who could be falsely classified as recently infected. McDougal et al. [63] used "specimens largely derived from early infection," which were known not to include individuals with AIDS symptoms or on ART treatment. Bärnighausen et al. [10] and Mermin et al. [73] did not apply inclusion criteria, but indicated that ART roll-out was not widespread at the time of the study.

Four studies used additional clinical information to reclassify or exclude individuals from the sample for BED assay testing. Three studies used a previous positive HIV test [35, 84, 18], two studies used ART status [35, 18], and two studies used AIDS diagnosis [48, 35]. Buchacz et al. [18] excluded individuals from their sample who were both classified as recent by the BED assay and identified as having long-standing infection by additional information. Priddy et al. [84] and Jiang et al. [48] excluded all individuals identified as having long-standing infection by additional information, independent of BED assay test results. Hall et al. [35] did not exclude any individuals from the sample but classified all individuals as non-recent who were identified as having long-standing infection, independent of their BED assay test results. Li et al. [58] collected information on ART and AIDS diagnosis, but found

that none of their study participants needed to be excluded or reclassified on the basis of this information. None of the other studies reported using extra clinical information for reclassification or exclusion from the sample for BED testing.

In addition to the two strategies described above to correct for BED assay imperfection, a few studies adjusted for selective HIV study participation. An important example is the study of Hall et al. [35] which used a method developed by Karon et al. [53] to estimate the annual number of recently infected individuals in the US based on the number of recently infected cases detected by the national case-reporting surveillance. This estimate was calculated by dividing the number of cases detected in the surveillance by an estimate of the probability of detection. We have not reviewed the remainder of these methods in detail because they are specific in their application to certain study designs rather than to the BED assay.

Finally, two approaches were used to deal with HIV-positive individuals that had missing information on recent infection (e.g. as a result of insufficient sample to allow BED assay testing after an initial HIV test). The first approach [18, 34] excludes these individuals from the incidence estimation sample. The second approach [52, 73, 78] assumes that the proportion of recent infections in these individuals is the same as the proportion in HIV-infected individuals with known recent infection status, and adjusts the incidence estimate accordingly.

Optical Density Cut-off and Calibration Parameters

Twenty-three of the thirty-nine studies reported optical density cut-offs, with seventeen using a value of 0.8 [111, 48, 58, 49, 5, 94, 40, 52, 73, 92, 10, 101, 35, 84, 63, 18, 34] and six using a value of 1.0 [44, 88, 113, 104, 87, 78]. One study used a value of 0.75 in addition to a value of 1.0 [104].

Window periods used in the calculation of incidence varied from 153 to 187 days. Of the nineteen conference abstracts, eleven [89, 96, 103, 110, 82, 104, 55, 102, 107, 108, 109] did not report the window period used, while four reported using a value of 153 days [114, 49, 5, 94], two reported 155 days [111, 101] and two reported 180 days [113, 19]. The window periods reported in the full papers are presented in Table 7.2. To compute annual incidence, unit consistency demands that a window period specified in days must be converted to units of years before being used in the calculation of incidence. In Table 7.2 we also report the length of year factor (365 or 365.25) used in each study and the corresponding effective window period specified in units of years.

In addition to the window period, studies using a second-generation formula required estimates of sensitivity, short-term specificity and long-term specificity, while studies using third-generation formulae only required estimates of the false-

recent rate. With the exception of one study [10] that used a locally-valid long-term specificity, all the studies that used a second-generation formula used the sensitivity, short-term specificity and long-term specificity as reported by McDougal et al. [63]. Some studies did, however, use varying degrees of precision for these parameters as reported in Table 7.2. All three studies using third-generation approaches estimated a false-recent rate for the local setting where the BED assay was applied [40, 52, 10], although one of them (Karita et al. [52]) did not use this estimate when calculating incidence.

Confidence Interval Calculation

With the exception of Wasinrapee et al. [114] who computed CIs using a normal approximation based only on the number of BED recent classifications (hereafter BED recent counts), none of the conference abstracts indicated how CIs were calculated. Two of the full-text articles [58, 92] did not report how CIs were calculated, while eight [48, 52, 73, 87, 121, 84, 63, 18] used a normal approximation based only on the number of BED recent counts, one [124] used a log transform of a normal approximation based only on BED recent counts, two [86, 93] used a normal approximation based on the BED recent and HIV-negative counts, one [78] used a modified Wald method, two [40, 35] used a delta method, and one [10] used a full trinomial distribution to approximate CIs. In all of these studies only uncertainty resulting from counting error was taken into account, while the remaining three papers [44, 88, 34] additionally accounted for the uncertainty associated with the window period by using a Bonferroni procedure that combined the window period CI with the CI for the counting error that was calculated using a Poisson distribution [47]. None of the studies using second- or third-generation approaches attempted to account for uncertainty stemming from error in the estimation of sensitivity and specificity parameters.

7.3.1 Incidence Formulae

We identified ten incidence formulae and classified them into three generations (Table 7.3), as described above (Figure 7.1). There were four first-generation formulae, two second-generation formulae and four third-generation formulae. The differences among the first-generation formulae stem from the different heuristics used to estimate the at-risk population, resulting in different denominators. The denominators of the two second-generation formulae differ for similar reasons. The reasons for differences in the third-generation formulae have been systematically explored elsewhere [69].

First-Generation Formulae

$$I = \frac{R}{N\omega + R} \tag{1}$$

$$I = \frac{R}{N\omega + R/2} \tag{2}$$

$$I = \frac{R}{(N+R)\omega} \tag{3}$$

$$I = \frac{R}{(N+R/2)\omega} \tag{4}$$

N is the number of HIV negative individuals,

P is the number of HIV positive individuals, and

R is the number of assay recent individuals.

Second-Generation Formulae

$$I = \frac{fR}{N\omega + fR} \qquad \text{McDougal et al. [63]} \tag{5}$$

$$I = \frac{fR}{N\omega + fR/2} \qquad \text{CDC [23]} \tag{6}$$

where

$$f = \frac{R/P + \rho_2 - 1}{(\sigma - \rho_1 + 2\rho_2 - 1)}.$$

Third-Generation Formulae

$$I = \frac{R - \varepsilon P}{R - \varepsilon P + (1-\varepsilon)N\omega} \qquad \text{Welte et al. [117]} \tag{7}$$

$$I = \frac{R - \varepsilon P}{R - \varepsilon(N+P) + N\omega} \qquad \text{Hargrove et al. [40]} \tag{8}$$

$$I = \frac{R - \varepsilon P}{R/2 - \varepsilon(N + P/2) + N\omega} \qquad \text{CDC [23]} \tag{9}$$

$$I_r = \frac{R - \varepsilon P}{(1-\varepsilon)N\omega} \qquad \text{McWalter/Welte [71]} \tag{10}$$

where $\varepsilon = 1 - \rho_2$. To convert form a rate (I_r) to an annual risk of infection (I) use $I = 1 - e^{-I_r}$.

Tab. 7.3: Formulae.

7.3.2 Sensitivity Analysis: Recalculation of Incidence Values

For all the full-text articles that unambiguously reported the information required, we recalculated HIV incidence to examine the impact of methodological choices. Table 7.4 shows the incidence estimates as computed by all formulae listed in Table 7.3. With the exception of a single entry, which may differ as a result of a rounding error, we were able to recover all the estimates reported in the original papers.

Where the original paper used only a first-generation approach, we evaluated second-generation formulae using the parameters of McDougal et al. [63] ($\sigma = 0.723$, $\rho_1 = 0.768$, $\rho_2 = 0.944$), and third-generation formulae using $\varepsilon = 1 - \rho_2 = 0.056$. Note that for studies using a window period other than 153 days, the use of these parameter values is inconsistent, because they were calibrated under the assumption of a 153-day window period.

Considerable variability in incidence estimates occurs due to the choice of estimator (Table 7.4). As expected, first-generation approaches produced larger incidence estimates when compared with second- and third-generation approaches. If the studies that used either a second- or third-generation approach had instead used a first-generation approach, their incidence estimates would have increased by between 9% (Li et al. [58], data for the year 2005) and 70% (Hargrove et al. [40]). Had the studies that used a first-generation approach used a second- or third-generation approach (with calibration parameters of McDougal et al. [63]), their incidence estimates would have decreased by between 27% (Priddy et al. [84]) and 97% (Jiang et al. [48], data for "City D All"). It is important to note that the calibration parameters used for determining these decreases are unlikely to be valid for the particular setting of each study; thus, the estimated decreases do not necessarily represent the true relative overestimation of HIV incidence. They do, however, emphasize the importance of estimator and parameter choices.

7.3.3 Sensitivity Analysis: Calibration Parameters

In order to explore the sensitivity of the incidence estimates to changes in the calibration parameters, we applied the McWalter/Welte estimator to the sample counts reported by full-text articles, using a range of false-recent rates and window periods. Across all studies, we determined the maximum and minimum parameter values used. To compute a conservative range of values for the sensitivity analyses, we added a margin of half the difference between maximum and minimum value to the maximum value (for both false recent rate and window period) and subtracted the same margin from the lowest value (only in the case of the window period, because a false-recent rate cannot be negative and the lowest observed value was zero). The

Authors	Reported incidence % (95% CI)	First generation				Second generation		Third generation				
		Formula 1	Formula 2	Formula 3	Formula 4	Formula 5	Formula 6	Formula 7	Formula 8	Formula 9	Formula 10	ARI
Bärnighausen et al.†	3.17 (2.54-3.80)	4.09	4.17	4.19	4.22	***0.66***	*0.66*	**3.12**	**3.19**	3.24	**3.22**	**3.17**
Buchacz et al.	9.5 (3.9-15.1)	8.82	9.23	9.29	**9.48**	*0.50*	*0.50*	*0.49*	*0.53*	*0.53*	*0.49*	*0.49*
Gupta et al.	1.0 (0.1-4.4)	1.02	**1.03**	1.03	1.03	*0.61*	*0.61*	*0.60*	*0.65*	*0.65*	*0.60*	*0.60*
Hargrove et al.	6.0 (5.2-6.9)	**9.55**	10.02	10.01	10.28	**5.49**	*5.64*	5.75	**6.04**	6.23	6.10	5.92
Hu et al.	17.3 (12.8-24.2)	15.79	17.14	**17.32**	18.01	*12.42*	*13.24*	*12.29*	*13.17*	*14.10*	*14.01*	*13.08*
Jiang et al. (City D All)	2.07 (1.26-2.89)	2.05	**2.07**	2.08	2.09	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*
Jiang et al. (City D IDU)	9.58 (5.83-13.34)	9.14	**9.58**	9.65	9.85	*3.98*	*4.06*	*3.93*	*4.26*	*4.35*	*4.09*	*4.01*
Karita et al. (Masaka)	7.5 (5.2-9.9)	7.24	**7.51**	7.56	7.68	6.15	**6.35**	6.09	6.60	**6.82**	6.48	6.28
Karita et al. (Kakira)	6.2 (4.4-8.0)	6.01	**6.20**	6.23	6.31	5.04	**5.17**	4.99	5.41	**5.56**	5.25	5.11
Li et al. (Data for 2005)‡	2.9 (0.8-5.0)	3.14	3.19	3.19	3.22	2.91	**2.95**	2.88	3.12	3.17	2.97	2.92
Li et al. (Data for 2006)	3.6 (1.3-5.9)	3.95	4.03	4.04	4.08	3.57	**3.63**	3.53	3.82	3.90	3.66	3.59
Mermin et al.	1.8 (1.5-2.1)	2.42	2.45	2.45	2.47	1.74	**1.76**	1.72	1.87	1.89	1.75	1.74
Nesheim et al.	0.24 (0.20-0.29)	**0.24**	0.24	0.24	0.24	*0.10*	*0.10*	*0.10*	*0.10*	*0.10*	*0.10*	*0.10*
Priddy et al.	1.3 (0.6-2.1)	1.32	1.33	**1.33**	1.34	*0.98*	*0.99*	*0.97*	*1.06*	*1.06*	*0.98*	*0.98*
Xiao et al. (IDUs)	8.2 (5.9-11.5)	7.90	**8.23**	8.29	8.43	*5.47*	*5.63*	*5.41*	*5.87*	*6.05*	*5.72*	*5.56*
Xiao et al. (County B)	4.2 (3.5-5.0)	4.09	**4.10**	4.19	4.23	*2.68*	*2.72*	*2.65*	*2.88*	*2.92*	*2.72*	*2.68*

The significant digits in the reported incidence column are the same as reported in each of the studies. The table includes only studies which provide sufficient information for independent incidence estimation. Underlined values indicate the primary incidence values recovered and bold-only values indicate the secondary incidence values recovered. Italics in second- and third-generation formulae indicate that values were calculated using parameters as reported in McDougal et al. (to 3 decimal places).

Key: ARI indicates annual risk of infection (using Formula 10 and conversion formula, see Table 3).
† Formula 5 yields 3.032% (as reported) when 0.9831 is used for long-term specificity.
‡ Possible rounding error on primary incidence.

Tab. 7.4: Incidence estimates.

resulting ranges were 0% to 8.4% for the false-recent rate and 136 to 204 days for the window period.

Figure 7.4 demonstrates the impact on incidence estimates of changing the false-recent rate. The figure shows the expected inverse relationship between incidence and false-recent rate. It also shows that the rate of change of incidence as a function of false-recent rate varies from survey to survey. This finding can be explained by the fact that the various populations surveyed have a ratio of recent infections to long-term infections that is relatively larger or smaller, suggesting a stage of epidemic that is more or less mature. For some cases, when the false-recent rate is too large, negative incidence values occur. This is due to the fact that the number of false-recent results is overestimated, with the consequence that the numerator in the estimator (Formula (10)) becomes negative.

Figure 7.5 demonstrates the impact on incidence estimates of changing the window period. Again, in all cases, incidence declines with increasing value of window period, but because the window period appears in the denominator, the declines are not linear. For the ranges of parameter values displayed, the declines are generally smaller than those that occur due to changes in the false-recent rate.

7.3.4 Recalculation of Confidence Intervals

In Table 7.5 we compute 95% CIs for the reproduced incidence estimates under several scenarios. To incorporate the effect of parameter uncertainty, we used a closed-form expression for the coefficient of variation of the estimator derived using a delta-method approximation [71]. Where a first-generation approach was used in the original study, we compute the annual risk of infection and uncertainty assuming that the false-recent rate is set to zero. In the column labeled "Counting error only", we report the CIs that result from counting uncertainty, excluding parameter uncertainty. In subsequent columns, we report CIs taking into account several values of the coefficient of variation for both the window period and the false-recent rate (where it is not zero).

7.4 Discussion

We have identified thirty-nine English-language articles published between 2003 and 2009 that used BED assay surveys to estimate HIV incidence. The use of the assay has generally increased since 2003 (the slight observed decline in the number of studies from fourteen in 2007 to eleven in 2008 may be due to delayed entry of studies into the PubMed database). Despite the increase in the use of the assay, the methods of its application have not converged to one approach.

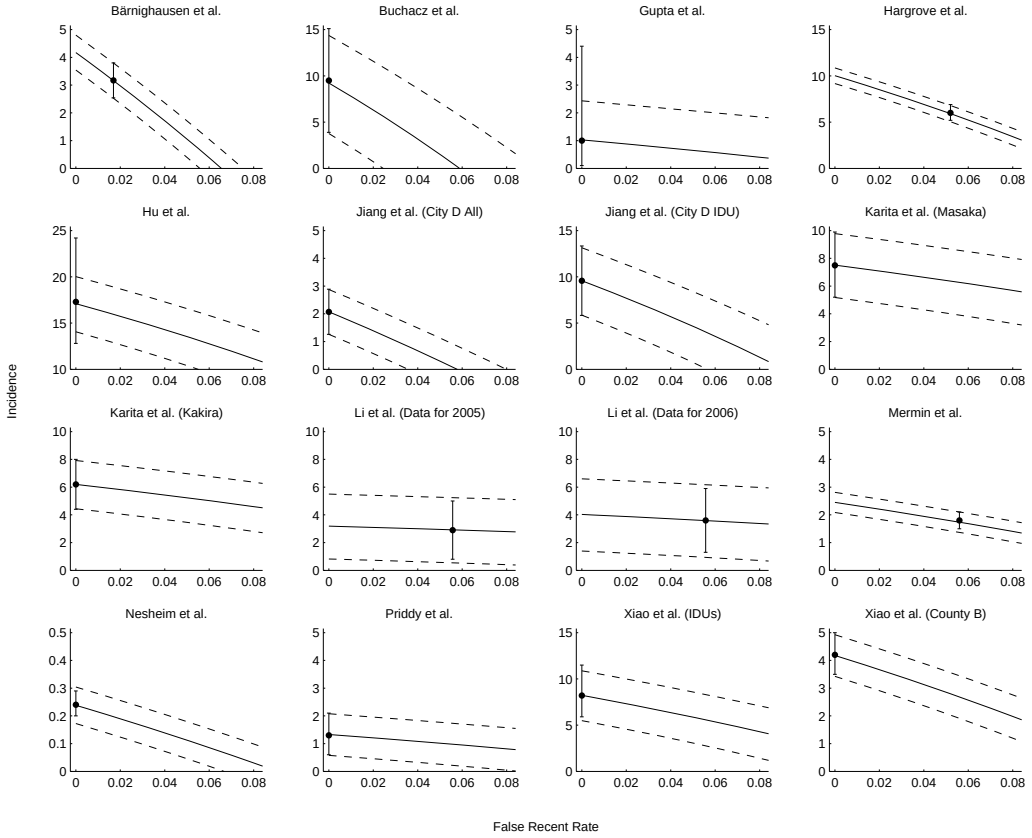


Fig. 7.4: Sensitivity of incidence estimates to changes in the false-recent rate (FRR). Filled circles and error bars show originally published estimates, solid lines show annual risk of infection as computed using Formula (10) and dashed lines show 95% confidence intervals due to counting error as determined using a delta method approximation [71] (excluding parameter uncertainty).

7.4.1 Estimator Choices

With three exceptions [10, 40, 52], all the studies reviewed used either a first-generation estimator (which assumes that the BED assay has a false-recent rate of zero, or a second-generation estimator (which adjusts for the fact that the BED assay is an imperfect test using three additional calibration parameters). The choice of method can produce very different incidence values. In particular, use of a second- or third-generation approach, with applicable calibration parameters, leads to substantially lower incidence estimates than those calculated using a first-generation approach.

All three studies [10, 40, 52] that used third-generation estimators implemented the approach of Hargrove et al. [40] and only one study implemented the estimator

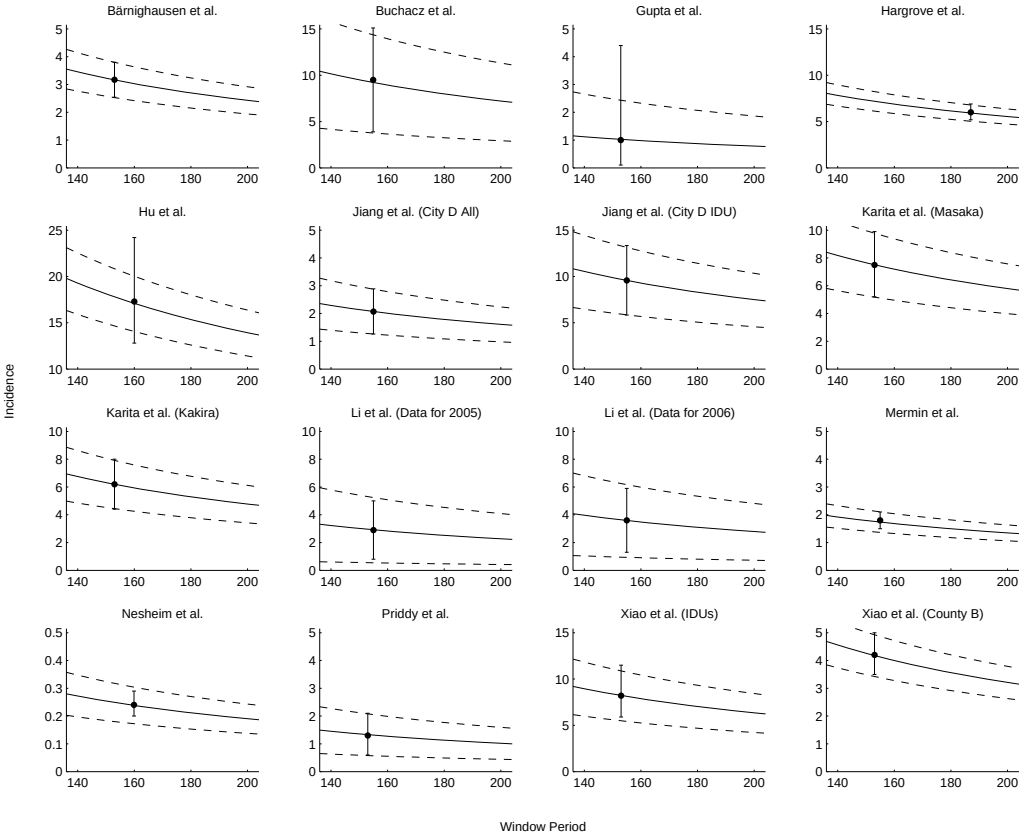


Fig. 7.5: Sensitivity of incidence estimates to changes in the window period. Filled circles and error bars show originally published estimates, solid lines show annual risk of infection and dashed lines show 95% confidence intervals (excluding parameter uncertainty).

Authors	Reported incidence % (95% CI)	ARI %	Counting error 95% CI	CoV 5% 95% CI	CoV 10% 95% CI	CoV 15% 95% CI	CoV 20% 95% CI
Bärnighausen et al.	3.17 (2.54-3.80)	3.17	2.53-3.80	2.45-3.87	2.26-4.06	2.00-4.31	1.72-4.59
Buchacz et al.	9.5 (3.9-15.1)	9.22	3.77-14.36	3.69-14.43	3.48-14.62	3.13-14.92	2.68-15.32
Gupta et al.	1.0 (0.1-4.4)	1.03	0.00-2.43	0.00-2.44	0.00-2.45	0.00-2.46	0.00-2.49
Hargrove et al.	6.0 (5.2-6.9)	5.92	5.04-6.79	4.78-7.04	4.24-7.57	3.59-8.19	2.90-8.84
Hu et al.	17.3 (12.8-24.2)	17.10	14.06-20.02	13.68-20.38	12.72-21.25	11.46-22.38	10.03-23.61
Jiang et al. (City D All)	2.07 (1.26-2.89)	2.07	1.26-2.88	1.24-2.90	1.17-2.97	1.06-3.08	0.93-3.21
Jiang et al. (City D IDU)	9.58 (5.83-13.34)	9.57	5.86-13.14	5.74-13.25	5.43-13.54	4.94-13.98	4.33-14.53
Karita et al. (Masaka)	7.5 (5.2-9.9)	7.51	5.18-9.78	5.07-9.89	4.77-10.17	4.32-10.59	3.79-11.09
Karita et al. (Kakira)	6.2 (4.4-8.0)	6.20	4.44-7.92	4.34-8.01	4.07-8.27	3.69-8.64	3.23-9.08
Li et al. (Data for 2005)	2.9 (0.8-5.0)	2.92	0.55-5.24	0.53-5.26	0.48-5.31	0.39-5.39	0.28-5.49
Li et al. (Data for 2006)	3.6 (1.3-5.9)	3.59	0.94-6.17	0.91-6.20	0.84-6.26	0.73-6.37	0.58-6.51
Mermin et al.	1.8 (1.5-2.1)	1.74	1.37-2.10	1.32-2.15	1.21-2.26	1.07-2.40	0.91-2.56
Nesheim et al.	0.24 (0.20-0.29)	0.24	0.17-0.30	0.17-0.31	0.16-0.32	0.14-0.33	0.12-0.35
Priddy et al.	1.3 (0.6-2.1)	1.33	0.58-2.08	0.57-2.09	0.53-2.12	0.48-2.17	0.42-2.24
Xiao et al. (IDUs)	8.2 (5.9-11.5)	8.23	5.49-10.88	5.38-10.99	5.07-11.28	4.60-11.71	4.03-12.23
Xiao et al. (County B)	4.2 (3.5-5.0)	4.18	3.43-4.93	3.32-5.03	3.07-5.27	2.75-5.59	2.39-5.93

The significant digits in the reported incidence column are the same as reported in each of the studies.
The table includes only studies which provide sufficient information for independent confidence interval estimation.

Key: ARI indicates annual risk of infection (using Formula 10 and conversion formula, see Table 3); CoV, coefficient of variation for parameters.

Tab. 7.5: Confidence intervals.

proposed by McWalter and Welte [71]. It is perhaps not surprising that third-generation estimators had more limited application because these estimators have only been recently published [40, 117, 71].

7.4.2 Calibration

It is incontrovertible that the BED assay produces a certain proportion of false-recent results. All studies that have estimated a false-recent rate for the BED assay have reported non-zero values [10, 63, 40, 52]. Three studies using a first-generation approach reported incidence estimates based on the BED assay that were substantially higher than incidence estimates using other techniques [44, 109, 113]. These results are consistent with the fact that failure to account for false-recent samples produces estimates that are too high. Two further studies reported that the BED assay overestimated incidence, but did not report the approach used to calculate incidence [19, 55]. Since first-generation approaches do not account for false-recent results, incidence estimates using these approaches should be considered invalid unless further justification for using a false-recent rate of zero can be provided. None of the studies reviewed provided such justification.

Second- or third-generation approaches provide more accurate incidence estimates if the correct calibration parameters are used. However, calibration parameters differ by setting, as illustrated by four of the studies reviewed. The original estimates for sensitivity and specificity were provided by McDougal et al. [63], who verified that a second-generation incidence estimate was consistent with an incidence estimate obtained in a longitudinal cohort study. Hargrove et al. [40] found that the false-recent rate for postpartum women in Zimbabwe is similar to that found by McDougal et al. [63], but that the mean window period was larger. Bärnighausen et al. [10] found that the false-recent rate for a population in rural KwaZulu Natal, South Africa, was lower than the value reported by Hargrove et al. and McDougal et al. All three of these studies found that BED assay-based incidence estimates, computed using local parameters, were similar to incidence estimates based on longitudinal data from the same population. Karita et al. [52] showed that the BED assay in Uganda using second- and third-generation approaches (with the parameters of McDougal et al. [63]) overestimates incidence values. They also reported prospective data that indicated a false-recent rate of 27% (8 of 30 individuals with follow-up data past one year) with a 95% CI of 12-46%, which is higher than the false-recent rate of 5.6% used in their incidence calculations. The binomial CI was not reported in the original paper; it was computed using the exact method. Unfortunately, this very uncertain point estimate for the false-recent rate leads to negative incidence estimates. If, however, a false-recent rate of 22% had been used, the BED-based

incidence estimates would have been similar to the values obtained by longitudinal surveillance (e.g. applying Formula (10) to their survey counts for Masaka gives an incidence of 1.5%, which is similar to the longitudinal incidence estimates of 1.7% and 1.3%).

With one exception, all studies that used the second generation approach did so in conjunction with the parameter values reported by McDougal et al. [63]. None of these studies independently calibrated sensitivity and specificity, or justified on other grounds why these values were likely to be correct for the particular study setting. This potentially undermines their validity. It is perhaps not surprising that researchers using second-generation approaches did so with parameter values from another setting, because studies to calibrate these parameters require frequent follow-up of large cohorts for periods of a year or longer, making calibration logistically complex and expensive to conduct [63].

It is important to emphasize that the use of exclusion or reclassification criteria should be applied consistently in the calibration-parameter-estimation and incidence-estimation surveys. Even if the value of a calibration parameter is locally estimated, it may not be appropriate for incidence estimation if the estimation sample differs systematically from the calibration sample. For example, McDougal et al. estimated long-term specificity using "specimens from longer-term-infected individuals not known to have clinical AIDS, opportunistic infections, or to be on treatment" [63]. The same exclusion criteria should thus be applied to samples in studies using the McDougal parameter values in incidence estimation. However, only one of the studies that used the McDougal parameter values collected information on AIDS diagnosis and antiretroviral treatment with the intention to exclude individuals from the sample for incidence calculation [58]. This issue of systematic differences between calibration and incidence samples is especially important if the differences relate to variables associated with the probability of false-recent BED assay diagnosis.

7.4.3 Confidence-interval Calculation

Comparing the CIs reported in the literature with those we computed under the assumption of no parameter uncertainty shows that the reported CIs were reasonable and in some cases overestimated uncertainty [44, 34]. Only three studies computed CIs that took into account the uncertainty of the calibration parameters [44, 34, 88]. When we included calibration uncertainty, the CIs of many incidence point estimates were substantially widened. We used somewhat artificial values for the coefficient of variation for the parameters—obviously, in real-life applications, the uncertainty due to error in the measurement of parameters will be a function of the statistical

power of the particular calibration study.

As shown by our CI calculations, surveys with fewer than several thousand individuals produce results with large uncertainty. Findings of studies with such small sample sizes should be interpreted with caution. We do not discuss the issue of sample size calculations further, but refer the reader to ongoing work on characterizing the necessary sample sizes to ensure reasonable precision [118].

7.5 Conclusion

Valid tests for recent HIV infection hold great promise for HIV research in that they allow the estimation of incidence using cross-sectional surveys. Such tests could thus substantially increase the capacity to monitor and understand the development of the epidemic and the impact of interventions at the population level. In the past five years, the BED assay has found worldwide application as a test for recent infection. However, as this review and sensitivity analysis demonstrate, many of the BED-derived HIV incidence estimates may not be valid. In particular, incidence estimates derived using first-generation approaches should be considered invalid, because they assume a false-recent rate equal to zero. Incidence estimates derived using second and third-generation approaches may be valid, but only if the calibration parameters are locally appropriate. Confidence intervals in general underestimated the associated incidence uncertainty because they did not account for parameter uncertainty.

It is possible to produce accurate incidence estimates if false-recent results are correctly accounted for and if studies recruit a sufficiently large number of participants. Based on these findings, we make recommendations for the future use of the BED assay. These recommendations are applicable not only to the BED assay, but also more broadly to other tests for recent HIV infection (including algorithms [56] with multiple tests and clinical information).

- Studies should report all information necessary for readers to independently determine incidence point and uncertainty estimates. In particular, studies should report (1) sufficient data to permit reconstruction of the population counts of HIV-negative, HIV-positive and recently-infected individuals; (2) the approach, formulae and parameter values used in incidence point estimation; and (3) the method used to estimate CIs.

- The estimation of incidence should be based on methods that account for false-recent results. In particular, the third-generation approach of McWalter and Welte [71] should be used because it is both parsimonious (as opposed to

second-generation estimators that are over-parameterized) and the least biased of the estimators reviewed [69].

- Locally-valid estimates of the calibration parameters should be used for computing incidence. If locally-valid estimates are not available, sensitivity of incidence to changes in these parameter values should be explored. It is particularly important to use an accurate estimate for the false-recent rate.

- As far as possible, further clinical or biomarker information should be used to reduce the false-recent rate. Furthermore, any additional information used to exclude or reclassify individuals should be applied consistently in both the study estimating the false-recent rate and the subsequent incidence estimation study. (For example, it would be incorrect to calculate incidence in a sample that includes individuals on ART using a false-recent rate estimated in a sample that excluded individuals on ART).

- CIs for incidence estimates should be computed using approaches that take into account parameter uncertainty.

Debate on the most appropriate approach for dealing with false-recent results continues [15, 116, 14, 39, 62]. The above recommendations reflect our judgments on best current practice as identified through a systematic review of published studies and theoretical considerations.

Chapter 8

Reply to 'Should biomarker estimates of HIV incidence be adjusted?'

∗ This chapter was coauthored with A. Welte and T. Bärnighausen [116], and is reproduced with permission from Wolters Kluwer Health (Rightslink licence no. 2601230822792): AIDS (2009) 23:2062-3 DOI: 10.1097/QAD.0b013e32832eff59.

8.1 Correspondence

Brookmeyer [15] is right to attempt the important task of reviewing and contrasting different approaches to biomarker-based HIV incidence estimates. The two 'results' highlighted in his abstract are as follows:

1. "The McDougal adjustment has no net effect on the estimate of HIV incidence because false positives exactly counterbalance false negatives".

2. "The Hargrove adjustment has a mathematical error that can cause significant underestimation of HIV incidence rates".

These findings appear to undermine the progress made in explaining why earlier BED assay-based methods have tended to overestimate incidence. However, both of Brookmeyer's results are incorrect. Given the evidence for subpopulations who fail to progress out of the biomarker-defined 'recent' category (so-called assay nonprogressors), 'adjustment' is indeed necessary.

Brookmeyer outlines a conception of incidence estimation requiring demographic and epidemic equilibrium conditions over the past M years, in which M is the maximum time an individual remains classified 'recent' by the biomarker. He then claims that M is 3 years for the BED assay, thus excluding the possibility of assay nonprogressors. This seems hard to sustain in light of various data of which we are aware [63, 40, 10]. Hargrove et al. [40] provide data on postpartum mothers indicating that 5.2% of those surveyed remain persistently classified as 'recent' by the BED assay. McDougal et al. [63] infer from their data that an individual has

a 5.6% probability (reported as a long-term specificity $\rho_2 = 0.994$) of testing below BED threshold, if infected longer than twice the mean window period of the assay. Under the assumption of no assay nonprogressors, Brookmeyer presents an argument to demonstrate that no 'adjustment' is required. His first result (point one above) is therefore inappropriate, as it depends on an assumption that is inconsistent with the data-driven findings in the publications he critiques.

Brookmeyer reports a numerical simulation in which the Hargrove estimator (using $\varepsilon = 0.052 = 1 - \rho_2$) apparently produces egregious underestimates of incidence, possibly even negative values. The cause of the underestimate is inconsistent calibration. The simulated epidemic has no assay nonprogressors, but he uses Hargrove's 'adjusted' estimator that assumes them to be 5.2% of the population. Although it is not reported, a near identical underestimate arises with the McDougal formula (when, equivalently, $\rho_2 = 0.948$ is used). The bias merely reflects that the incidence estimators are unavoidably very sensitive to the calibration of ρ_2, a very important and usually neglected point [10]. Conversely, if one samples or simulates a population in which there is a subpopulation of assay nonprogressors, then 'unadjusted' estimators are well known to overestimate incidence because a disproportionate number of 'false recent' classifications accumulate in the population. In this situation, the McDougal and Hargrove estimators, when appropriately calibrated, yield results with modest bias, dominated by counting error for reasonable sample sizes [40, 10]. We provide an analytical closed-form demonstration of inherent bias in each of these methods [69]. Brookmeyer's other result (point two above), thus incorrectly attributes substantial bias exclusively to the Hargrove estimator when in fact both the McDougal and Hargrove estimators exhibit similar bias, which results from Brookmeyer's inconsistent calibration of the estimator.

As we have shown elsewhere [69, 117], it is possible to simplify the McDougal framework, under its own assumptions, but not as Hargrove or Brookmeyer suggest. Detailed analysis reveals an identity relating the sensitivity and specificity parameters, leading to a simpler estimator that is easier to calibrate. We have also derived a formally rigorous framework for biomarker-based incidence estimation that specifically accounts for assay nonprogressors [71], and can also account for assay regressors under suitable calibration [10]. This approach requires fewer assumptions and is less prone to bias than either the McDougal or Hargrove method.

Brookmeyer notes the unsatisfactory correspondence between published biomarker-based incidence estimates and estimates based on prospective follow-up. His discussion of possible sources of error focuses on sampling bias and imperfect mean window period estimation. Although these issues are important, he proposes no way of dealing with assay nonprogressors. In his conclusion, Brookmeyer remarks

that "if, however, a proportion of HIV-positive persons are identified who remain in the window period indefinitely, then an adjustment would be necessary". This does little to soften his strong statements, which undermine prior work addressing the issue of nonprogressors that has helped us move beyond the naive estimators.

Using data from cross-sectional surveys to estimate incidence will remain an attractive approach, but it requires the use of a robust estimator for which the correct applicable calibrations have been performed. In particular, accurate calibration of long-term specificity is of vital importance to correctly account for biomarker misclassification.

Chapter 9

Incidence from Cross-sectional Surveys: Improved Characterization of Tests for Recent Infection

∗ This chapter was coauthored with A. Welte and R. Kassanjee [65].

Abstract

Background: Since it is cheaper, quicker and easier than prospective follow-up, incidence estimation from cross-sectional surveys has gained much attention. The estimators used with this methodology require a characterization of the Test for Recent Infection (TRI), which has variously been specified using a combination of mean window periods, sensitivities, specificities and false-recent rates. Recent research has highlighted problems with such characterizations and raised debate about how best to specify TRI properties.

Methods: By introducing a predetermined cutoff time (τ), and making the assumption of constant incidence for a period τ preceding the survey, we provide a precise and parsimonious characterization of a TRI, and derive a new incidence estimator. The estimator depends on three parameters; the probability of remaining in the window period at a time τ after infection, the mean window period for those who leave it before τ and the proportion of the subpopulation infected for longer than τ that are (incorrectly) classified recent by the test. The new estimator is contrasted with the previous estimators of McDougal et al., Hargrove et al. and McWalter & Welte.

Results: Although the epidemiological assumptions required are more restrictive than those of the McWalter/Welte method, the characterization of the TRI is more general, better defined and more amenable to estimation. The McDougal and McWalter/Welte estimators are shown to be special cases of the new estimator. The extent to which the previous estimators differ from the new estimator is shown under various scenarios. The assumption of constant incidence for a period τ before the survey is shown to be benign under realistic epidemic scenarios.

Conclusion: With a precise formulation of the TRI parameters and the as-

sociated incidence estimator, the issues that have caused debate in the area of cross-sectional incidence estimation are clarified.

9.1 Introduction

It is of considerable epidemiological importance to develop methodologies for estimating disease incidence. For some time there has been substantial interest in methods for estimating incidence from cross-sectional data [17, 47, 80, 63, 40, 71]. Prevalence, in general, is a reflection of historical incidence, convolved with the susceptible population dynamics and survival after disease acquisition. This convolution renders disease prevalence an indirect and unresponsive proxy for recent incidence, unless disease duration is very short. In the case of lifelong infections such as HIV, a more direct view of recent incidence is required. Cross-sectional testing for "recent infection" in principle provides a simple proxy for recent incidence, but tests for "recent infection", based on some combination of laboratory assays and clinical information exhibit considerable inter-subject variability, including subpopulations with anomalous responses, sometimes loosely referred to as *false-recent results*. These lead to subtle complications in the interpretation of data and derivation of provably consistent estimators, which have raised some controversy in the recent literature [15, 39, 62, 116, 14]. This article briefly reviews the topic and presents new analysis which synthesizes the ideas currently in contention.

A generic framework for recent infection tests applied to HIV can be formulated by considering an immune response which increases[1] over time post infection. A quantitative result (typically an optical density of some analyte) can be cast into a categorical result by use of a classification threshold, below which the response is regarded as indicative of "recent infection". An application of a cross-sectional survey classifies individuals of the sample population as *healthy* (susceptible) individuals, or infected individuals who are either *under* or *over* the specified threshold on the immune system response. The counts in the survey, corresponding to these categories, are denoted N_{H}, N_{U} and N_{O} respectively.

Developers of recent infection tests face a fundamental trade-off between the need for the duration of recent infection to be long enough, to ensure that the population proportion of individuals classified as recent is sufficiently large to provide statistical power in sample populations of attainable size, and not so long that the test is unrepresentative of "recent infection". For example, a duration of one month is too short, while a duration of three years is too long.

All candidate biomarkers exhibit non-trivial inter-subject variability in response,

[1] There is no loss of generality if the response decreases with time.

which is a source of complication in the analysis. Of serious concern is the subpopulation of individuals that exhibit anomalous responses on the assay. Two important classes are the assay non-progressors, who never develop enough of a response to HIV infection to cross the threshold, and the regressors, who initially cross the threshold but then drop below it again (e.g. due to immune failure as a result of late-stage disease progression, or due to treatment). When the classification threshold for a biomarker is increased there is a corresponding increase in the duration of recent infection. There is, however, also an increase in the proportion of anomalous results. Thus, developers must also face another fundamental trade-off between length of duration and the proportion of anomalous results.

We review a number of approaches that have been suggested to deal with these anomalous cases and describe some of the shortcomings associated with each. Initially, we describe a simple estimator that does not directly account for anomalous results. We then describe generalizations of this approach that have been suggested in the literature.

All these methods require some investment in parameterizing the interaction of the proposed test with the intended study population (i.e., estimator calibration). There is currently no consensus on a unified framework within which to compare these approaches or to address their limitations [15, 39, 62, 116, 14]. By providing precise definitions for the parameters that characterize a test for recent infection, we derive a new estimator which unifies the previous approaches.

9.2 Previous Estimators

A recent infection test in which all individuals progress to the recent/non-recent infection threshold, and there is no significant reversion below it, leads to the simple estimator

$$I = \frac{N_{\mathrm{U}}}{\omega N_{\mathrm{H}}}, \tag{9.1}$$

where ω is the duration of recency (also known as the mean window period). In the limit of a slowly varying susceptible population, this provides an estimate of the weighted incidence, with the natural weighting proportional to the availability for being infected and remaining classified recent once infected [51, 71]. The incidence weighting scheme is given explicitly by

$$I_{\mathrm{W}} = \frac{\int_{-\infty}^{0} I(t)H(t)S_{\mathrm{U}}(-t)\,dt}{\int_{-\infty}^{0} H(t)S_{\mathrm{U}}(-t)\,dt},$$

where $I(t)$ is instantaneous incidence, $H(t)$ is the susceptible population and $S_{\mathrm{U}}(t)$ is survival in the state of being under-threshold on the assay (i.e., being classified

as recent). Alternatively, the incidence estimated in this manner can be interpreted as the "incidence in the recent past" [16].

This estimator may be used even when anomalous results are present—there are, however, two undesirable consequences. Firstly, incidence will be weighted over the time periods for which the anomalous results are present. When assay non-progressors are present, this means that the estimated incidence is a weighted average over the whole course of the epidemic. This effect has been demonstrated in a recent paper by Brookmeyer [16], where he calculates a "shadow time" associated with such weighted incidence estimates. These shadow times are unreasonably large when even a small percentage of assay non-progressors are present in the population of interest. Secondly, and perhaps more seriously, the calibration of ω becomes more onerous. When a subpopulation of assay non-progessors is present, then estimation of ω must take place over the full life-span of these individuals. Moreover, since individuals in this subpopulation exit the "recently-infected" category as a result of death or treatment, this means that ω is less a property of the assay and more dependent on environmental factors that are likely to change with location and time (e.g. access to primary health care, ART coverage, local mix of opportunistic challenges and nutrition). For example, this implies that ω in North America would be very different to ω in rural Africa. Since estimation of the mean window period is difficult and expensive, requiring follow-up over the lifetime of the non-progressors, this means that providing an accurate location- and time-specific estimate of ω for use with estimator (9.1) will be difficult. In particular, it is hardly realistic to expect that a calibration study be conducted over a period of 10 to 20 years, and that analysis of cross-sectional data should wait for completion of such a study. Moreover, since ω is time dependent, the result of such a calibration study may not be valid for very long. This implies that, if even a very small proportion of individuals are assay non-progressors, it will be impractical to calibrate this estimator.

As a result of these complications, a number of approaches have been proposed to estimate, and account for, the effect of anomalous test results. The first approach was proposed by McDougal and colleagues [63]. They impose hard time boundaries to define up to three states of infection (recent, non-recent and long), which are then conceived as imperfectly reflected in the test. An estimator making use of three extra parameters (sensitivity, short-term specificity and long-term specificity) is introduced.

Hargrove and colleagues [40], under the assumption that sensitivity is approximately equal to short-term specificity provide a reduced estimator that depends only on a mean window period and a false-recent rate, being one minus the long-term specificity in the McDougal approach. More recently, it has been shown that there is

a consistent method for reducing the parameters in the McDougal approach [117, 69]. The biases inherent in the McDougal and Hargrove approaches have also been investigated [69].

The reduction of the McDougal estimator and bias comparisons for the McDougal and Hargrove parameters were inspired and facilitated by the derivation of a new estimator [71] that makes fewer assumptions than either of these two approaches. In the next section we provide an intuitive derivation of this approach, but point the reader to the original paper for the full survival analysis derivation.

For the remainder of the paper, we shall refer to the following simple incidence estimator

$$I = \frac{N_{\mathrm{TR}}}{\omega N_{\mathrm{H}}}, \tag{9.2}$$

where N_{TR} is the estimate of the number of individuals in the sample that are "truly recent". The subtlety will be in defining precisely what "recency" means, and providing the procedures and parameters required to estimate N_{TR} and ω. By addressing these issues we aim to provide a new incidence estimator that accounts for the sources of false-recent results and has clear guidelines on how it should be calibrated.

9.3 Accounting for Assay Non-progressors

All previous attempts to derive incidence estimators, that account for false-recency, have made the assumptions of no regression and equal survival for assay progressors and assay non-progressors [69]. Under these assumptions, we have previously derived an incidence estimator using survival analysis [71]. Briefly, "recent infection" and the parameters that describe it are defined as follows:

- For individuals that progress on the assay, recency means testing below-threshold.
- The duration of recency ω_{R} is defined as the mean time spent under the threshold, for those individuals that progress.
- False-recent results are produced by individuals that fail to progress on the assay. The probability of not progressing is denoted by $\mathbb{P}_{\mathrm{NP}}$.

The first point above means that recent infection is a characteristic of the test and is not defined by a definite time boundary. If we now "assign" recency times for the non-progressors from the same distribution applicable to assay progressors then we

can define an explicit incidence weighting scheme as follows:

$$I_{\mathrm{W}} = \frac{\int_{-\infty}^{0} I(t)H(t)S_{\mathrm{R}}(-t)\,dt}{\int_{-\infty}^{0} H(t)S_{\mathrm{R}}(-t)\,dt},$$

where, as before, $I(t)$ is instantaneous incidence and $H(t)$ is the susceptible population. In contrast to the previous weighting scheme, $S_{\mathrm{R}}(t)$ is now survival in the state of being under-threshold on the assay *conditional on progressing.* In the limit of a slowly varying susceptible population, the denominator may be written as $\omega_{\mathrm{R}} N_{\mathrm{H}}$ (compare with (9.2)), and so it remains to find an estimate of the numerator

$$N_{\mathrm{TR}} = \int_{-\infty}^{0} I(t)H(t)S_{\mathrm{R}}(-t)\,dt.$$

The assumptions of no regression and equal survival mean that, at every stage in the epidemic, the ratio of assay non-progressors to assay progressors is $\mathbb{P}_{\mathrm{NP}}$ to $1-\mathbb{P}_{\mathrm{NP}}$. This is also the ratio of the number of false-recent results (denoted N_{FR}) to the number of individuals with a test result over-threshold:

$$\frac{N_{\mathrm{FR}}}{N_{\mathrm{O}}} = \frac{\mathbb{P}_{\mathrm{NP}}}{1-\mathbb{P}_{\mathrm{NP}}}.$$

The number of true-recent results is the number of individuals classified by the biomarker as under the threshold minus the number of false-recent results:

$$N_{\mathrm{TR}} = N_{\mathrm{U}} - N_{\mathrm{FR}} = N_{\mathrm{U}} - \frac{\mathbb{P}_{\mathrm{NP}}}{1-\mathbb{P}_{\mathrm{NP}}} N_{\mathrm{O}}.$$

In conjunction with the expression for the denominator, this gives the estimator

$$I_1 = \frac{N_{\mathrm{U}} - \frac{\mathbb{P}_{\mathrm{NP}}}{1-\mathbb{P}_{\mathrm{NP}}} N_{\mathrm{O}}}{\omega_{\mathrm{R}} N_{\mathrm{H}}} = \frac{N_{\mathrm{U}} - \mathbb{P}_{\mathrm{NP}}(N_{\mathrm{U}} + N_{\mathrm{O}})}{\omega_{\mathrm{R}}(1-\mathbb{P}_{\mathrm{NP}})N_{\mathrm{H}}}. \tag{9.3}$$

The number of true-recent results in this estimator is the same as that which arises in the estimator of McDougal *et al.* [63]. Their estimator does, however, use a different denominator (See [69] for further details).

9.4 Relaxing Assumptions

By relaxing the assumptions of no regression and equal survival for assay progressors and assay non-progressors, an incidence estimator, similar to the previous one, may by obtained using a slightly different definition of the parameters that describe recency:

- For individuals that progress on the assay, recency means testing below-threshold.

- The duration of recency ω_{R} is defined as the mean time spent under the threshold, for those individuals that progress.

- False-recent results are produced by individuals that fail to progress on the assay and individuals that revert below the threshold as a result of immune failure or treatment. A general false-recent rate (FRR)[2], denoted by ϕ, is defined as the proportion of non-recent infections that are incorrectly classified as recent.

The first two points are the same as in the previous section, and, as a result, so is the incidence weighting scheme. The FRR described in the third point is a nontrivial function of the progression of the epidemic (including historical incidence and susceptible population), the survival functions for assay progressors and non-progressors, and the rate at which regression occurs.

To derive a new expression for N_{TR}, we note that the definition of the FRR may be written as

$$\phi = \frac{N_{\mathrm{FR}}}{N_{\mathrm{O}} + N_{\mathrm{FR}}},$$

which can be rearranged to provide an estimate of the number of false-recent results

$$N_{\mathrm{FR}} = \frac{\phi}{1-\phi} N_{\mathrm{O}}.$$

As before, the number of true-recent results is the difference between the number of individuals that are under-threshold and the number of false-recent results, i.e., $N_{\mathrm{TR}} = N_{\mathrm{U}} - N_{\mathrm{FR}}$. Thus, the estimator can be written as

$$I_2 = \frac{N_{\mathrm{U}} - \frac{\phi}{1-\phi} N_{\mathrm{O}}}{\omega_{\mathrm{R}} N_{\mathrm{H}}} = \frac{N_{\mathrm{U}} - \phi(N_{\mathrm{U}} + N_{\mathrm{O}})}{\omega_{\mathrm{R}}(1-\phi) N_{\mathrm{H}}}, \tag{9.4}$$

which is the same as (9.3), except that $\mathbb{P}_{\mathrm{NP}}$ has been replaced by ϕ.

These definitions raise the question, "Is it possible to specify a precise and consistent procedure to estimate the parameters (ω_{R} and ϕ) required by this estimator?"

If one is sure that all individuals that will progress on the assay do so before a known maximum progression time (T), then there are at least two consistent methods to provide estimates for ω_{R}. The first involves prospectively enrolling individuals with an approximately known infection time into a follow-up study (with frequent visits) for a period at least as long as the maximum progression time. An empirical survival curve of being under the threshold, for individuals that progress, is then

[2] Note that the FRR is in fact a false positive rate. We prefer to use the word 'recent' instead of 'positive' to emphasize that it relates to the test for recency, and not to the test for disease positivity.

produced and integrated to obtain an estimate of the mean duration[3]. The second method involves using a maximum likelihood approach on individuals that are followed-up at intervals greater than the maximum progression time [115, 54].

Estimates of ω_{R} will be biased if there are a significant number of late progressors, i.e., individuals that progress at a time post-infection longer than the assumed cutoff T. This raises the question, "How can one be confident that the value of T chosen is at least as large as the maximum progression time?" This is an issue that has been raised in the recent debate in the literature on biomarker based estimators [15, 39, 62, 116, 14].

The estimation of ϕ is perhaps more problematic. In a previous paper [10], a cutoff time T equal to twice the mean window period was chosen, and the FRR was estimated as the proportion of under-threshold individuals with a known infection time greater than T. This makes the assumption that all such individuals are falsely classified as recent, which may not be true if late progressors are present.

The FRR, as defined, is a product of the whole history of the epidemic. However, the procedure outlined in the last paragraph produces an estimate for the FRR that is only applicable to the subgroup of individuals with a time since infection greater than T. If the proportion of false-recent results for the subgroup of individuals with a time since infection less than T is different from this estimated FRR, then this procedure provides a biased estimate of ϕ. An unbiased estimate of ϕ is a non-trivial mixture of the FRR estimates applicable to both subgroups.

As noted previously, the change from using a probability of not progressing to using a general false-recent rate has not fundamentally changed the incidence weighting scheme described in the previous section. If the maximum progression time T is known, however, then one may reduce the limits of integration and the weighted incidence can be written as

$$I_{\mathrm{W}} = \frac{\int_{-T}^{0} I(t)H(t)S_{\mathrm{R}}(-t)\,dt}{\int_{-T}^{0} H(t)S_{\mathrm{R}}(-t)\,dt}.$$

9.5 New Definitions for Parameters and a New Estimator

The observation that knowledge of the maximum progression time changes the limits of integration on the incidence weighting scheme suggests a new scheme in which one specifies a cutoff time and explicitly writes the weighting scheme with the lower

[3] This is similar to the procedure used by McDougal et al. [63], except that it is not clear from their paper whether or not they removed the assay non-progressors when computing their curve.

limit of integration specified by the cutoff time. New definitions for the parameters that describe recency may then be proposed.

- For individuals that progress on the assay *before* a *prespecified* cutoff time τ, recency means testing below-threshold.
- The duration of recency ω_τ is defined as the mean time spent under the threshold, for those individuals that progress on the assay *before* τ.
- False-recent results are produced by the fraction of individuals that fail to progress on the assay *before* τ. The probability of not progressing before τ is denoted by $\mathbb{P}_\tau$.
- All individuals with an infection time longer than τ are by definition non-recent infections. An FRR ϕ_τ is defined as the proportion of individuals infected for longer than τ who are classified (falsely) as recent.

The first point means that recent infection is now a characteristic of the test *and* the cutoff time, which is conveniently chosen (e.g. one year). The last point provides a precise definition of an epidemic dependent FRR more amenable to estimation than the FRR proposed in the previous section. Note that the three parameters ω_τ, $\mathbb{P}_\tau$ and ϕ_τ have been written with the subscript τ to emphasize that they are functions of the value chosen for τ. Of course, one would like τ to be chosen so that most individuals who progress do so before that time. This ensures that the false-recent rate, as defined, will not unduly depend on incidence in the recent past.

The incidence weighting scheme is then written as

$$I_{\mathrm{W}} = \frac{\int_{-\tau}^{0} I(t)H(t)S_{\mathrm{R}}(-t)\,dt}{\int_{-\tau}^{0} H(t)S_{\mathrm{R}}(-t)\,dt}. \tag{9.5}$$

As before, in the limit of a slowly varying susceptible population, the denominator may be written as $\omega_\tau N_{\mathrm{H}}$, and so it remains to find an estimate of the numerator

$$N_{\mathrm{TR}} = \int_{-\tau}^{0} I(t)H(t)S_{\mathrm{R}}(-t)\,dt. \tag{9.6}$$

Let the number of HIV seropositive individuals in the population be given by n (i.e., $n = N_{\mathrm{U}} + N_{\mathrm{O}}$). The number of seropositive individuals can also be written as

$$n = \int_{-\infty}^{0} I(t)H(t)S(-t)\,dt,$$

which depends on the historical incidence, the number of susceptible individuals and the post-infection survival of HIV infected individuals $S(t)$. Denote the number of

seroconverters with a time since infection less than or equal to τ by $n_{t\leq\tau}$ and those with a time greater than τ by $n_{t>\tau}$. The number of individuals in the population with a time since infection less than τ can be written as

$$n_{t\leq\tau} = \int_{-\tau}^{0} I(t)H(t)S(-t)\,dt. \tag{9.7}$$

We now make the following three observations:

- The total number of individuals in the population that are under-threshold is equal to the number of individuals that are truly recent infections plus the number of false-recent results.
- The number of individuals with an infection time less than τ that are *truly* non-recent is $n_{t\leq\tau} - N_{\text{TR}}$. By definition, $\mathbb{P}_\tau$ specifies the fraction of these individuals that have failed to progress prior to τ, and, as a result, are incorrectly classified as recent. Hence the number of false-recent results with an infection time less than τ is $\mathbb{P}_\tau(n_{t\leq\tau} - N_{\text{TR}})$.
- The number of false-recent results with an infection time greater than τ is just the FRR, ϕ_τ, multiplied by $n_{t>\tau} = N_{\text{U}} + N_{\text{O}} - n_{t\leq\tau}$.

These observations lead to the expression

$$N_{\text{U}} = N_{\text{TR}} + \mathbb{P}_\tau(n_{t\leq\tau} - N_{\text{TR}}) + \phi_\tau(N_{\text{U}} + N_{\text{O}} - n_{t\leq\tau}). \tag{9.8}$$

Now, in order to get an estimate for $n_{t\leq\tau}$, make the assumption of uniform infection events in the population over the last time period τ, i.e., $I(t)H(t)S(-t) = f_0$. This is the same assumption made by McDougal et al. [63, 69] and is equivalent to the following assumptions:

- Incidence is constant over the last period τ,
- Susceptible population is constant over the last period τ, and
- No death as a result of infection occurs in a period τ post-infection (i.e., $S(t) = 1$ for $t \in [0, \tau]$).

These assumptions, in conjunction with (9.7) and (9.6), mean that

$$n_{t\leq\tau} = \tau f_0 = \frac{\tau}{\omega_\tau}(\omega_\tau f_0) = \frac{\tau}{\omega_\tau} N_{\text{TR}},$$

which, in conjunction with (9.8), yields an expression for N_{TR} given by

$$N_{\text{TR}} = \frac{N_{\text{U}} - \phi_\tau(N_{\text{U}} + N_{\text{O}})}{1 - \mathbb{P}_\tau + \frac{\tau}{\omega_\tau}(\mathbb{P}_\tau - \phi_\tau)}.$$

Having computed an estimate for the numerator (compare with (9.2)) we obtain a new estimator for incidence given by

$$I_3 = \frac{N_{\mathrm{U}} - \phi_\tau(N_{\mathrm{U}} + N_{\mathrm{O}})}{(\omega_\tau(1 - \mathbb{P}_\tau) + \tau(\mathbb{P}_\tau - \phi_\tau))\, N_{\mathrm{H}}}. \tag{9.9}$$

Note that if $\mathbb{P}_\tau = \phi_\tau = \mathbb{P}_{\mathrm{NP}}$ we recover the previous estimator (9.3), and if $\mathbb{P}_\tau = \phi_\tau = \phi$ we recover estimator (9.4).

As mentioned previously, the assumptions made in deriving this estimator are the same assumptions made in the derivation (and reduction [69]) of the McDougal estimator [63], with one important exception. McDougal et al. made the assumption that the empirical survival curve, of individuals in the state of recency, used for calibration was flat after $\tau = 2\omega$. This assumption has been relaxed by allowing $\mathbb{P}_\tau$ to be different from ϕ_τ.

9.6 Numerical Simulations

The assumption of constant incidence and susceptible population used in deriving the new estimator is stringent, and it is necessary to understand the kinds of biases that arise when realistic transient conditions occur. For this reason, we explore the bias that may be expected under realistic violations of this assumption.

For demonstration purposes, suppose incidence and susceptible population vary according to the exponential functions of time given by

$$I(t) = I_0 e^{\alpha t} \qquad \text{and} \qquad H(t) = H_0 e^{\beta t}, \tag{9.10}$$

where I_0 and H_0 are the instantaneous incidence and susceptible counts at the time of the survey, and α and β are factors expressing the rate at which incidence and the susceptible population are changing. We also assume that the survival function S_{R} is generated using a Weibull distribution with scale parameter $l = 0.44$ and shape parameter $k = 7$, corresponding to a mean of 150 days and a standard deviation of 25 days. A cutoff time $\tau = 1$ year is selected, which means that $\omega_\tau = 150$ days, and the probability of not progressing was set to $\mathbb{P}_\tau = 2\%$.

Figure 9.1 shows the bias, being the relative difference between I_{W} and I_3, as a function of the FRR ϕ_τ and the incidence growth rate (expressed as an annual percentage, i.e., -50% means a halving of incidence in one year, while 100% means a doubling). This plot was produced for the situation in which there is no growth of the susceptible population. See the Appendix for a derivation of the expressions used.

When the annual growth rate of the susceptible population is set to 10% the plot produced is the same as that shown, except the surface is shifted upwards

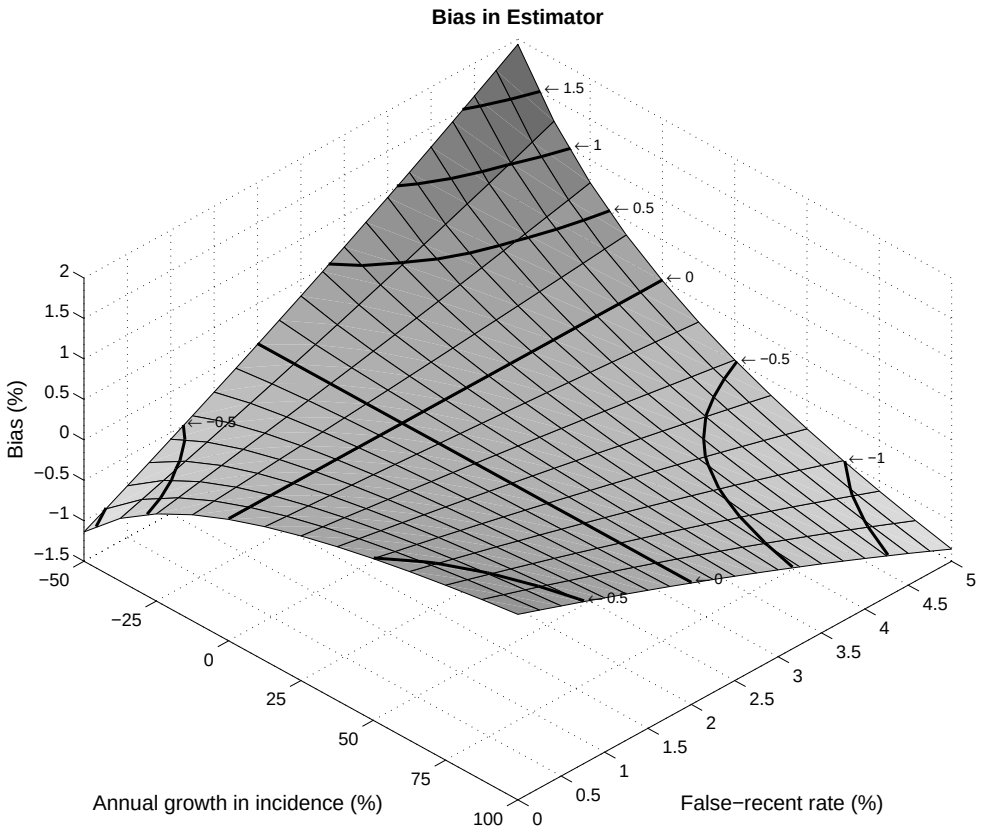


Fig. 9.1: Bias between I_{W} and I_3 expressed as a percentage.

by approximately 2%. Similarly, when the annual growth rate is set to -10% the surface is shifted downward by about 2%. This is consistent with the bias findings in McWalter & Welte [71].

This exercise shows that the bias introduced by making the assumption of constant incidence and susceptible population over a time period τ prior to the survey is benign—of the order of a few percent.

It is also interesting to explore the extent to which the estimator I_2 is different from I_3. Using the parameter choices $\omega_{\mathrm{R}} = \omega_\tau$ and $\phi = \phi_\tau$ for the estimator I_2, the Appendix derives an expression for the relative difference between the two estimators as a function of δ, being the relative difference between $\mathbb{P}_\tau$ and ϕ

$$\delta = \frac{\mathbb{P}_\tau - \phi}{\phi}.$$

Figure 9.2 shows the relative difference between the estimators as a function of ϕ_τ in the range 0 to 5% and the extent to which $\mathbb{P}_\tau$ is different from ϕ_τ expressed as a relative percentage difference (i.e., $\delta = -100\%$ means $\mathbb{P}_\tau = 0$ and $\delta = 100\%$ means that $\mathbb{P}_\tau = 2\phi_\tau$). The difference is also of the order of a few percent, but large

enough to warrant further investigation.

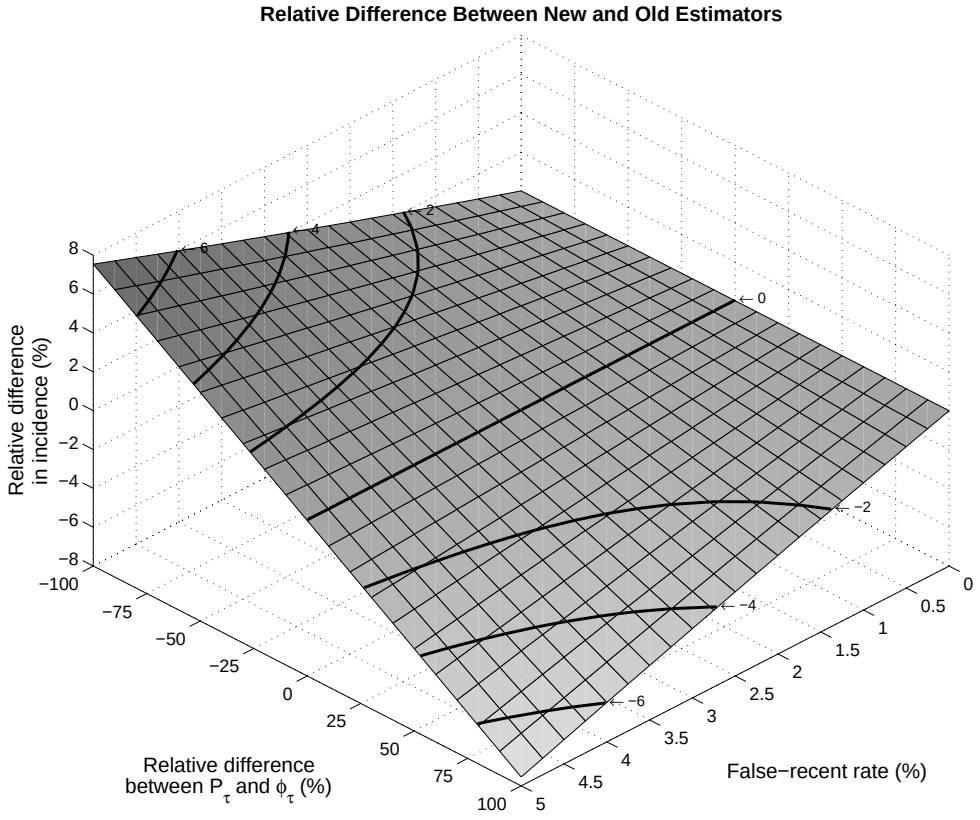


Fig. 9.2: Relative difference between I_3 and I_2 expressed as a percentage.

9.7 Appendix

9.7.1 Bias Computation

The expressions for $I(t)$ and $H(t)$ given by (9.10) mean that

$$I_{\mathrm{W}} = \frac{I_0 \int_{-\tau}^{0} e^{(\alpha+\beta)t} S_{\mathrm{R}}(-t)\, dt}{\int_{-\tau}^{0} e^{\beta t} S_{\mathrm{R}}(-t)\, dt}.$$

In order to evaluate I_3, recall that

$$N_{\mathrm{U}} = N_{\mathrm{TR}} + \mathbb{P}_\tau(n_{t\leq\tau} - N_{\mathrm{TR}}) + \phi_\tau(N_{\mathrm{U}} + N_{\mathrm{O}} - n_{t\leq\tau}),$$

in which case the numerator of I_3 can be written as

$$(1 - \mathbb{P}_\tau)N_{\mathrm{TR}} + (\mathbb{P}_\tau - \phi_\tau)n_{t\leq\tau}.$$

We also have

$$N_{\mathrm{TR}} = H_0 I_0 \int_{-\tau}^{0} e^{(\alpha+\beta)t} S_{\mathrm{R}}(-t)\, dt$$

and

$$n_{t\leq\tau} = H_0 I_0 \int_{-\tau}^{0} e^{(\alpha+\beta)t}\, dt.$$

Substituting these expressions into the estimator I_3 and noting that $N_{\mathrm{H}} = H_0$ yields

$$I_3 = I_0 \left(\frac{(1-\mathbb{P}_\tau)\int_{-\tau}^{0} e^{(\alpha+\beta)t} S_{\mathrm{R}}(-t)\, dt + (\mathbb{P}_\tau - \phi_\tau)\int_{-\tau}^{0} e^{(\alpha+\beta)t}\, dt}{\omega_\tau(1-\mathbb{P}_\tau) + \tau(\mathbb{P}_\tau - \phi_\tau)} \right).$$

The bias may now be computed using numerical integration on the expression for the relative difference

$$r = \frac{I_{\mathrm{W}} - I_3}{I_{\mathrm{W}}}.$$

Note that, when the relevant substitutions are performed, this expression is independent of I_0 and H_0, which means that the bias is only dependent on the rates of growth of the incidence and susceptible population, not the values at the time of the survey.

9.7.2 Relative Difference Between I_3 and I_2

Here, we derive an expression for the relative difference (r) between estimator I_3 and estimator I_2

$$r = \frac{I_3 - I_2}{I_3},$$

assuming that $\phi = \phi_\tau$ is used as the FRR when computing I_2. To simplify expressions we use $\omega = \omega_{\mathrm{R}} = \omega_\tau$ and express $\mathbb{P}_\tau = \phi + \delta\phi$ in terms of the relative difference (δ) between $\mathbb{P}_\tau$ and ϕ_τ. Then I_3 may be expressed as

$$I_3 = \frac{N_{\mathrm{U}} - \phi(N_{\mathrm{U}} + N_{\mathrm{O}})}{(\omega(1-\phi-\delta\phi) + \tau\delta\phi)N_{\mathrm{H}}},$$

with

$$r = 1 - \frac{N_{\mathrm{U}} - \phi(N_{\mathrm{U}} + N_{\mathrm{O}})}{\omega(1-\phi)N_{\mathrm{H}}} \times \frac{(\omega(1-\phi-\delta\phi) + \tau\delta\phi)N_{\mathrm{H}}}{N_{\mathrm{U}} - \phi(N_{\mathrm{U}} + N_{\mathrm{O}})} = \frac{(\omega-\tau)\delta\phi}{\omega(1-\phi)}.$$

Remarkably, the final expression is independent of the sample counts. This also means that the estimate I_3 can be written as a constant factor multiplied by the estimate I_2

$$I_3 = I_2 \left(\frac{\omega_{\mathrm{R}}(1-\phi)}{\omega_\tau(1-\mathbb{P}_\tau) + \tau(\mathbb{P}_\tau - \phi)} \right),$$

which can be deduced by directly inspecting (9.4) and (9.9).

Appendix A

Selected MATLAB Code

A.1 Fig2-5.m

```
%
% Program to simulate a population and test incidence estimator on a
% number of cross-sectional samples. This simulation uses Weibull
% survival functions.
%

T=50;               % Number of years
nsteps=50;          % Number of time steps

H0=100000;          % Initial number of susceptables
H1=H0/20;           % Linear rate of change of susceptables

Pnp=0.05;           % Probability of remaining in recently infected state
eps=Pnp/(1-Pnp);

lamle=8.83;         % Scale parameter for life expectancy
kle=4.5;            % Shape parameter for life expectancy
lamw=0.44;          % Scale parameter for window period
kw=7;               % Shape parameter for window period

ssize=5000;         % Sample size

rand('seed',1);

itime=zeros(1,1500000);
ni=0;
t=0;

while t<T
    ni=ni+1;

    if t<10         % Set parameters for linear hazard rate based on current time
        I1=0;
        I0=0.01;
    elseif t>=10 && t<20
        I1=0.009;
        I0=0.01-10*I1;
    elseif t<30
        I1=0;
        I0=0.1;
    elseif t<40
        I1=-0.007;
        I0=0.1-30*I1;
    else
        I1=0;
        I0=0.03;
    end

%   Use cubic roots formula

    z=cubic(I1*H1/3,(I0*H1+I1*H0)/2,I0*H0,log(rand)-I1*H1*t^3/3-(I0*H1+I1*H0)*t^2/2-I0*H0*t);
    z=z(imag(z)==0);
    t=min(z(z>t));

    itime(ni)=t;
end

itime=itime(1:ni);

%%

falserecents=rand(1,ni)<Pnp;                                % Determine which individuals are false-recent

longtime=itime+lamw.*(-log(rand(1,ni))).^(1/kw);            % Times are Weibull distributed
death=itime+lamle.*(-log(rand(1,ni))).^(1/kle);
```

```
omega=lamw*gamma(1+1/kw);                                       % Compute window period
omsqr=lamw^2*(gamma(1+2/kw)-gamma(1+1/kw)^2)+omega^2;

weightedinc=zeros(1,nsteps); sampleinc=zeros(1,nsteps); CoV=zeros(1,nsteps);
PH=zeros(1,nsteps); PU=zeros(1,nsteps); PO=zeros(1,nsteps);

for t=1:nsteps
    ctime=t*T/nsteps;
    Rt=sum(itime<ctime & longtime>ctime & death>ctime);         % True recents
    Rf=sum(longtime<ctime & death>ctime & falserecents);        % False recents
    L=sum(longtime<ctime & death>ctime & not(falserecents));    % Longs
    S=H0+floor(H1*ctime);                                       % Susceptables
    totpop=S+Rt+Rf+L;                                           % Total population

    PH(t)=S/totpop;                                             % Compute population proportions
    PU(t)=(Rt+Rf)/totpop;
    PO(t)=1-PH(t)-PU(t);

    samp=[];                                                    % Sample the population
    while length(samp)<ssize
        samp=[samp rand(1,ssize-length(samp))*totpop];
        samp=unique(ceil(samp));                                % Ensure sampling without replacement
    end

    NH=sum(samp<=(totpop-(Rt+Rf+L)));                           % Compute sample counts
    NU=sum(samp<=(totpop-L))-NH;
    NO=ssize-NH-NU;

    weightedinc(t)=Rt/(S*omega-H1*omsqr/2);                     % Incidence using linear formula and true recents
    sampleinc(t)=(NU-eps*NO)/(omega*NH);                        % Sample incidence (simple formula)

    % Compute coefficent of variation

    CoV(t)=sqrt((1/PH(t)+(PO(t)*PU(t)*(1+eps)^2)/(PU(t)-PO(t)*eps)^2)/(ssize*(PO(t)+PU(t))));
end

%%

fprintf('Time %d   PH %.5f   PU %.5f   PO %.5f\n',15,PH(15),PU(15),PO(15));
fprintf('Time %d   PH %.5f   PU %.5f   PO %.5f\n',20,PH(20),PU(20),PO(20));
fprintf('Time %d   PH %.5f   PU %.5f   PO %.5f\n',30,PH(30),PU(30),PO(30));
fprintf('Time %d   PH %.5f   PU %.5f   PO %.5f\n',35,PH(35),PU(35),PO(35));
fprintf('Time %d   PH %.5f   PU %.5f   PO %.5f\n',40,PH(40),PU(40),PO(40));
fprintf('Time %d   PH %.5f   PU %.5f   PO %.5f\n',50,PH(50),PU(50),PO(50));

t=(1:nsteps)/nsteps*T;
close all;
figure('Position',[1,1,1600,700]);
axes('FontSize',12);

I1=0;
I0=0.01;
plot(0:10,I0+(0:10).*I1,':k');

hold;

plot(t,weightedinc,'k');
plot(t,sampleinc,'+k');
plot(t,weightedinc+weightedinc.*2.*CoV,'-.k');

I1=0.009;
I0=0.01-10*I1;
plot(10:20,I0+(10:20).*I1,':k');

I1=0;
I0=0.1;
plot(20:30,I0+(20:30).*I1,':k');

I1=-0.007;
I0=0.1-30*I1;
plot(30:40,I0+(30:40).*I1,':k');

I1=0;
I0=0.03;
plot(40:51,I0+(40:51).*I1,':k');

plot(t,weightedinc-weightedinc.*2.*CoV,'-.k');

title('Monte Carlo Experiment','fontsize',14,'fontweight','b');
xlabel('Years','fontsize',14);
ylabel('Incidence','fontsize',14);
axis([0 51 0 0.14]);
set(gcf,'color','none');

legend('Instantaneous incidence','Weighted incidence (total population)','Estimated incidence (sample)', ...
       'Two standard deviation envelope',2);
```

A.2 Cubic.m

```
function z=cubic(a,b,c,d)

%
% Function to compute roots of a cubic equation. Used by Fig2-5.m
%

if a~=0
    a1=b/a;
    a2=c/a;
    a3=d/a;
    Q=(3*a2-a1^2)/9;
    R=(9*a1*a2-27*a3-2*a1^3)/54;
    D=Q^3+R^2;
    Dsqrt=sqrt(D);
    if D<0
        S=(R+Dsqrt)^(1/3);
        T=conj(S);
    else
        RpD=R+Dsqrt;
        RmD=R-Dsqrt;
        if RpD<0 S=-(-RpD)^(1/3); else S=RpD^(1/3); end    % Make sure we get real cube roots!
        if RmD<0 T=-(-RmD)^(1/3); else T=RmD^(1/3); end
    end
    sqrt3=sqrt(3);
    z=[S+T -(S+T)/2+i*sqrt3*(S-T)/2 -(S+T)/2-i*sqrt3*(S-T)/2]-a1/3;
elseif b~=0
    D=sqrt(c^2-4*b*d);
    z=[(-c-D)/(2*b) (-c+D)/(2*b)];
else
    z=-d/c;
end
return
```

A.3 Fig4-2.m

```
%
% Compute the probability of failing to detect an incidence reduction when
% in fact the incidence halves.
%

Pnp=0.05;                   % Probability of not progressing
eps=Pnp./(1-Pnp);
omega=0.425;                % Window period
Inc=0.005:0.0025:0.05;      % Incidence range
alpha=11;                   % Average survival time once infected
Ifactor=.5;                 % Factor indicating change in incidence
N1=5000:1250:25000;         % Number of samples for survey 1
N2=N1;                      % Number of samples for survey 2
signif=0.05;                % Alpha

% Functions for incidence estimator and CoV
Est=@(H,U,O,Pnp,omega) (U-O*Pnp/(1-Pnp))/(H*omega);
CoV=@(PH,PU,PO,N,Pnp) sqrt((1/PH+(PO*PU*(1+Pnp/(1-Pnp))^2)/(PU-PO*Pnp/(1-Pnp))^2)/(N*(PO+PU)));

% Functions to compute steady state proportions
k=@(eps,omega,alpha) eps+(1+eps)*omega/(alpha-omega);
PU=@(I,eps,omega,alpha) omega*I*k(eps,omega,alpha)/(k(eps,omega,alpha)-eps+omega*I*(1+k(eps,omega,alpha)));
PO=@(U,eps,omega,alpha) U/(eps+(1+eps)*omega/(alpha-omega));

Prob=zeros(length(N1),length(Inc));

for I=1:length(Inc)
    for J=1:length(N1)

        % Compute population proportions for two surveys
        PU1=PU(Inc(I),eps,omega,alpha);
        PO1=PO(PU1,eps,omega,alpha);
        PH1=1-PU1-PO1;
        Prev=PU1+PO1;
        PU2=(Ifactor*Inc(I)*omega*(1-Prev)+eps*Prev)/(1+eps);
        PO2=Prev-PU2;
        PH2=1-PU2-PO2;

        % Compute standard deviations
        SDev1=CoV(PH1,PU1,PO1,N1(J),Pnp)*Est(PH1,PU1,PO1,Pnp,omega);
        SDev2=CoV(PH2,PU2,PO2,N2(J),Pnp)*Est(PH2,PU2,PO2,Pnp,omega);

        % Compute Null hypothesis population proportions
        PHN=(N1(J)*PH1+N2(J)*PH2)/(N1(J)+N2(J));
        PUN=(N1(J)*PU1+N2(J)*PU2)/(N1(J)+N2(J));
        PON=(N1(J)*PO1+N2(J)*PO2)/(N1(J)+N2(J));

        % Compute probability of failing to detect decrease
        SDevN1=CoV(PHN,PUN,PON,N1(J),Pnp)*Est(PHN,PUN,PON,Pnp,omega);
        SDevN2=CoV(PHN,PUN,PON,N2(J),Pnp)*Est(PHN,PUN,PON,Pnp,omega);
```

```
        SDevReal=sqrt(SDev1^2+SDev2^2);
        SDevNull=sqrt(SDevN1^2+SDevN2^2);

        Ilim=-sqrt(2)*erfcinv(2*(1-signif/2))*SDevNull;
        meandiff=Est(PH1,PU1,PO1,Pnp,omega)-Est(PH2,PU2,PO2,Pnp,omega);

        Prob(J,I)=1-0.5*(1+erf((Ilim-meandiff)/(SDevReal*sqrt(2))));
    end
end

%%

% Graph probability

close all;

figure('Position',[1,1,1200,1000],'Colormap',[(256:-1:0)'./512+0.4 (256:-1:0)'./512+0.4 (256:-1:0)'./512+0.4]);
axes('FontSize',12);
surf(Inc.*100,N1./1000,Prob*100);
xlabel(sprintf('Initial incidence\n (per 100 pyar)'),'Fontsize',14,'Position',[-22.41 -109.488 286.624]);
ylabel(sprintf('Sample size\n(thousands)'),'Fontsize',14,'Position',[-25.216 -97.015 286.624]);
zlabel('Probability of detecting reduction (%)','Fontsize',14);
axis([min(Inc)*100 max(Inc)*100 min(N1)/1000 max(N1)/1000 0 100]);
set(gca,'xtick',.5:.5:5);
set(gca,'ytick',5:2.5:25);
view([315 20]);
title(sprintf('Probability of Detecting Reduction in Incidence\n
    (Final incidence=half of initial incidence, \\omega=%.0f days, \\epsilon=%.0f%%, \\alpha=%.0f%%)',
    omega*365,Pnp*100,signif*100),'Fontsize',14,'fontweight','b');

% Compute and display contour

hold on;
set(gcf,'DefaultLineLineWidth',2);
C=contour3(Inc.*100,N1./1000,Prob*100+0.1,[90+0.1 90+0.1],'-k');
text(C(1,end),C(2,end),90+0.2,' \leftarrow 90','HorizontalAlignment','left','fontsize',10)

% Produce a 2D contour plot

figure('Position',[1,1,1200,1000]);
axes('FontSize',15);
[C,h]=contour(Inc.*100,N1./1000,Prob*100,[0.5 1 2.5 5 10 20 30 40 50 60 70 80 90 95 99],'k','ShowText','on');
text_handle = clabel(C,h);
set(text_handle,'Fontsize',14);
title(sprintf('Probability of Detecting Incidence Reduction - Contours of Constant Probability (%%)\n
    (Final incidence=half of initial incidence, \\omega=%.0f days, \\epsilon=%.0f%%, \\alpha=%.0f%%)',
    omega*365,Pnp*100,signif*100),'Fontsize',20,'fontweight','b');
xlabel(sprintf('Initial incidence (per 100 pyar)'),'Fontsize',18);
ylabel(sprintf('Sample size (thousands)'),'Fontsize',18);
set(gca,'xtick',.5:.5:5);
set(gca,'ytick',5:2.5:25);
```

A.4 Fig9-1.m

```
%
% Compute bias for new estimator using exponential incidence and
% susceptable population
%

lamw=0.44;                                  % Scale parameter for window period
kw=7;                                       % Shape parameter for window period
omega=lamw*gamma(1+1/kw);

inc=@(t,alpha) exp(log(1+alpha/100).*t);    % Exponential incidence
sus=@(t,beta) exp(log(1+beta/100).*t);      % Exponential susceptibles

Pnp=0.02;                                   % Probability of not progressing
tau=1;                                      % Cutoff time
Phirange=0:0.0025:0.05;                     % Range of FRRs
alpharange=-50:12.5:100;                    % Range of incidence growth rates
beta=0;

bias=zeros(length(Phirange),length(alpharange));

for i=1:length(alpharange)

    % Compute integrals numerically

    truedenom=quad(@(x)exp(-(x./lamw).^kw).*sus(-x,beta),0,tau);
    Ntr=quad(@(x)exp(-(x./lamw).^kw).*inc(-x,alpharange(i)).*sus(-x,beta),0,tau);
    HIVposletau=quad(@(x) inc(-x,alpharange(i)).*sus(-x,beta),0,tau);

    weightedInc=Ntr/truedenom;

    % Compute bais
```

```
    for j=1:length(Phirange)
        Phi=Phirange(j);
        estInc=((1-Pnp)*Ntr+(Pnp-Phi)*HIVposletau)/(omega*(1-Pnp)+tau*(Pnp-Phi));
        bias(j,i)=(weightedInc-estInc)/weightedInc;
    end
end

%%

% Graph computed bias

close all;

figure('Position',[1,1,1200,1000],'Colormap',[(256:-1:0)'./512+0.4 (256:-1:0)'./512+0.4 (256:-1:0)'./512+0.4]);
axes('FontSize',12);
surf(alpharange,Phirange*100,bias*100);
view([45 50]);
axis([-50 100 0 5 -1.5 2]);
set(gca,'xtick',-50:25:100);
set(gca,'ytick',0:0.5:5);
title('Bias in Estimator','fontsize',14,'fontweight','b');
xlabel('Annual growth in incidence (%)','fontsize',14,'Position',[660.289 -20.241 21.578]);
ylabel('False-recent rate (%)','fontsize',14,'Position',[711.902 -18.52 21.578]);
zlabel('Bias (%)','fontsize',14);

% Compute and display contours

hold on;
set(gcf,'DefaultLineLineWidth',2);
el=0.01;
el2=0.05;
[alpha,Phi]=meshgrid(alpharange,Phirange);
C=contour3(alpharange,Phirange*100,alpha*100+el,[el el],'-k');
text(C(1,2),C(2,2),0+el2,' \leftarrow 0','HorizontalAlignment','left','fontsize',10)
C=contour3(alpharange,Phirange*100,(Phi-Pnp)*100+el,[el el],'-k');
text(C(1,end),C(2,end),0+el2,' \leftarrow 0','HorizontalAlignment','left','fontsize',10)
C=contour3(alpharange,Phirange*100,bias*100+el,[0.5+el 0.5+el],'-k');
text(C(1,2),C(2,2),.5+el2,' \leftarrow 0.5','HorizontalAlignment','left','fontsize',10)
text(C(1,end),C(2,end),.5+el2,' \leftarrow 0.5','HorizontalAlignment','left','fontsize',10)
C=contour3(alpharange,Phirange*100,bias*100+el,[1+el 1+el],'-k');
text(C(1,2),C(2,2),1+el2,' \leftarrow 1','HorizontalAlignment','left','fontsize',10)
C=contour3(alpharange,Phirange*100,bias*100+el,[1.5+el 1.5+el],'-k');
text(C(1,2),C(2,2),1.5+el2,' \leftarrow 1.5','HorizontalAlignment','left','fontsize',10)
C=contour3(alpharange,Phirange*100,bias*100+el,[-0.5+el -0.5+el],'-k');
text(C(1,C(2,1)+3),C(2,C(2,1)+3),-0.5+el2,' \leftarrow -0.5','HorizontalAlignment','left','fontsize',10)
text(C(1,C(2,1)+1),C(2,C(2,1)+1),-0.5+el2,' \leftarrow -0.5','HorizontalAlignment','left','fontsize',10)
C=contour3(alpharange,Phirange*100,bias*100+el,[-1+el -1+el],'-k');
text(C(1,C(2,1)+3),C(2,C(2,1)+3),-1+el2,' \leftarrow -1','HorizontalAlignment','left','fontsize',10)
```

Appendix B

Poster Submitted to IAS Conference, Cape Town, 2009 [70]

A New Paradigm for Incidence Estimation From Cross-sectional Data

Thomas A. McWalter and Alex Welte (University of the Witwatersrand, Johannesburg, and SACEMA)

Background

Incidence measurement using cross-sectional surveys is an attractive approach because it is cheaper, quicker, potentially less-biased and logistically simpler to implement than prospective follow-up. The basic approach is summarized as follows:

- Test the HIV status of a representative cohort. This yields two counts: N the number of HIV negative and P the number HIV positive individuals.
- Among HIV positive individuals, find the number R of 'recent infections' as classified by a biomarker.
- Under ideal circumstances, the survey counts can be used to compute incidence using the estimator

$$I = \frac{R}{N\omega},$$

where the window period ω is the mean time individuals spend classified as 'recent' by the biomarker.

Unfortunately, biomarkers like the BED assay [6] are known to produce false positive results (i.e. non-recently infected individuals that are classified as recent). Controversy has arisen over how to account for these false positive results.

Old Paradigm

Under the assumption of epidemic equilibrium, McDougal et al. [3] proposed an approach to correct for false positive results which may be summarized as follows:

- Define 'recent infection' to mean an infection for a period shorter than ω.
- False positives are individuals that test recent but have been infected for longer than the window period ω. False negatives are individuals that are infected for shorter than ω but produce a non-recent biomarker result.
- Estimate the sensitivity σ, short-term specificity ρ_1 (infected between ω and 2ω) and long-term specificity ρ_2 (infected longer than 2ω) for the assay.
- Using survey counts, compute incidence using the estimator

$$I = \frac{fR}{N\omega + fR}, \qquad \text{where} \qquad f = \frac{R/P + \rho_2 - 1}{\sigma - \rho_1 + 2\rho_2 - 1}.$$

Calibration of this approach is complex, and since it is difficult to consistently incorporate calibration uncertainty into confidence intervals for incidence estimates this has never been done.

New Paradigm

Without the need for epidemic equilibrium, we have derived a new incidence estimator using survival analysis [5]. Under the special case of equilibrium, the approach may be summarized as follows:

- Identify the proportion ε of individuals that will never be classified as non-recent by the test — these are called assay non-progressors.
- For individuals that progress on the assay, define 'recent infection' to mean classified as recent by the biomarker. Recent infection is now a characteristic of the test, not a definite time boundary. The window period ω is now defined as the mean time assay progressors spend classified as recent.
- False positives are assay non-progressors with a time since infection larger than ω. There are no false-negatives under this definition.
- Using survey counts, compute incidence using the estimator

$$I = \frac{R - \varepsilon P}{(1-\varepsilon)N\omega}.$$

The model can be generalized to the case where individuals regress, i.e. revert to being classified as recently infected by the test as a result of end stage AIDS or use of ARVs. In this case, the parameter ε is interpreted as a false positive rate and must be appropriately calibrated for the setting in which the survey takes place.

Confidence Intervals

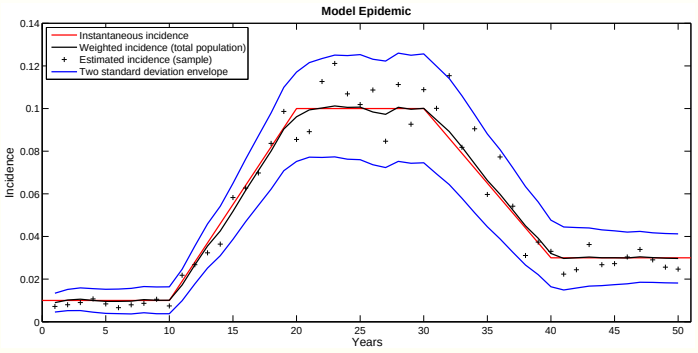
Confidence intervals for the estimator are derived using the delta method. They include the error associated with calibration parameters, specified as normal distributions with standard deviations σ_ε and σ_ω. The coefficient of variation of incidence is:

$$C_v = \sqrt{\frac{1}{P}\left(\frac{N+P}{N} + \frac{(P-R)R[1+\varepsilon/(1-\varepsilon)]^2}{[R-\varepsilon/(1-\varepsilon)(P-R)]^2}\right) + \frac{\sigma_\omega^2}{\omega^2} + \frac{\sigma_\varepsilon^2(P-R)^2}{(1-\varepsilon)^4[R-\varepsilon/(1-\varepsilon)(P-R)]^2}},$$

with 95% confidence intervals computed as $I \pm 1.96 \times C_v I$.

Estimator Validation

A model epidemic was simulated in which infection times were generated using a non-homogeneous Poisson process, with individual recency times and lifetimes generated using Weibull distributions. The susceptible population was linearly increasing with $S(t) = 100,000 + 5,000t$ (t in years). The red curve shown in the graph below is the model instantaneous incidence parameter used. The target for the estimates was the realized weighted incidence (black line), which was calculated for the complete population. This is flanked by a two standard deviation counting error envelope (blue lines). Simulated incidence estimates (+ symbols) were obtained by drawing cross-sectional samples of 5,000 individuals from the simulated population. Calibration parameters $\omega = 150$ days and $\varepsilon = 5\%$ were assumed to be known exactly.



The above simulation shows that even under demographic and epidemic non-equilibrium conditions the estimator performs well. Obviously, more precise incidence estimates can be obtained with larger sample sizes.

We have also compared the performance of the estimator applied to field data with an incidence estimate obtained from prospective follow-up [1], with favorable results. A key point arising from this validation exercise is that a locally applicable calibration for the false positive rate ε must be used.

Parameter Reduction and Bias Comparison

Under the assumption of a steady state epidemic, we have shown [7] that there is a relationship between the sensitivity and specificity parameters of the McDougal approach:

$$\sigma - \rho_1 + \rho_2 = 1.$$

This allows the elimination of σ and ρ_1 from their estimator. Since $\rho_2 = 1 - \varepsilon$, it also means that the parameters that remain are those required in the new paradigm.

Hargrove et al. [2] have proposed another estimator which results from a similar reduction of the McDougal estimator under the assumption that $\sigma = \rho_1$.

Comparing these two estimators with the new estimator, under a model steady state epidemic, only the new estimator recovers an unbiased estimate of the incidence [4].

Results

The key findings of this work may be summarized as follows:

- The new estimator is applicable under more relaxed assumptions than any of the previous estimators [5].
- Robust results are achieved when applied to a model epidemic which includes significant transients.
- Confidence intervals are accurate under model epidemic conditions.
- When compared to the other estimators, the new estimator is least biased [4].
- When applied, under consistent calibration, to field data it performs well compared to the incidence found by prospective follow-up [1].
- The calibration parameters ω and ε may vary with setting and care should be taken to ensure that valid estimates for the particular setting are used [1].

Conclusion

This work explores some of the issues that have lead to the controversy surrounding incidence measurement from cross-sectional surveys. In particular, it presents a consistent estimator with accurate confidence bounds. It also highlights the crucial role of using locally valid estimates for calibration parameters.

References

[1] Bärnighausen et al. HIV incidence in rural South Africa: Comparison of estimates from longitudinal surveillance and cross-sectional cBED assay testing. *PLoS ONE*, 3(11):e3640, 2008.

[2] J.W. Hargrove et al. Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. *AIDS*, 22(4):511–518, 2008.

[3] J.S. McDougal et al. Comparison of HIV type 1 incidence observed during longitudinal follow-up with incidence estimated by cross-sectional analysis using the BED capture enzyme immunoassay. *AIDS Research and Human Retroviruses*, 22(10):945–952, 2006.

[4] T.A. McWalter and A. Welte. A comparison of biomarker based incidence estimators. *Unpublished*, 2009.

[5] T.A. McWalter and A. Welte. Relating recent infection prevalence to incidence with a sub-population of assay non-progressors. *Journal of Mathematical Biology*, In press, 2009.

[6] B.S. Parekh et al. Quantitative detection of increasing HIV type 1 antibodies after seroconversion: A simple assay for detecting recent HIV infection and estimating incidence. *AIDS Research and Human Retroviruses*, 18(4):295–307, 2002.

[7] A. Welte, T.A. McWalter, and T. Bärnighausen. A simplified formula for inferring HIV incidence from cross-sectional surveys using a test for recent infection. *AIDS Research and Human Retroviruses*, 25(1):125–126, 2009.

Bibliography

[1] 16th Conference on Retroviruses and Opportunistic Infections: visit past conferences, Available at: `http://www.retroconference.org/2009/display.asp?page=181`, 2009.

[2] Africa Centre for Health and Population Studies, Available: `http://www.africacentre.ac.za/`.

[3] A. Agresti, *Categorical data analysis*, 2nd ed., Wiley, 2002.

[4] R.N. Anderson and H.M. Rosenberg, *Age standardization of deaths rates: implementation of the year 2000 standard*, National Vital Statistics Reports **47** (1998), 1–20.

[5] T. Apornpong, T. Pankum, S. Sirivichayakul, N. Phanuphak, P. Cunningham, E. Stein, J. Kaldor, K. Page-Shafer, and P. Phanupak, *HIV incidence using BED testing among clients attending anonymous HIV testing site in Bangkok, Thailand*, 4th IAS Conference on HIV Pathogenesis and Treatment and Prevention, Sydney, Australia, 2007.

[6] F. Barin, L. Meyer, R. Lancar, C. Deveau, M. Gharib, A. Laporte, J.C. Desenclos, and D. Costagliola, *Development and validation of an immunoassay for identification of recent human immunodeficiency virus type 1 infections and its use on dried serum spots*, Journal of Clinical Microbiology **43** (2005), no. 9, 4441–7.

[7] T. Bärnighausen, V. Hosegood, I.M. Timaeus, and M.-L. Newell, *The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa*, AIDS **21** (2007), no. Suppl 7, S29–S38.

[8] T. Bärnighausen, T.A. McWalter, Z. Rosner, M.-L. Newell, and A. Welte, *HIV incidence estimation using the BED capture enzyme immunoassay: Systematic review and sensitivity analysis*, Epidemiology **21** (2010), no. 5, 685–697, (Reproduced in this document as Chapter 7).

[9] T. Bärnighausen, F. Tanser, Z. Gqwede, C. Mbizana, K. Herbst, and M.-L. Newell, *High HIV incidence in a community with high HIV prevalence in rural South Africa: findings from a prospective population-based study*, AIDS **22** (2008), 139–144.

[10] T. Bärnighausen, C. Wallrauch, A. Welte, T.A. McWalter, N. Mbizana, J. Viljoen, N. Graham, F. Tanser, A. Puren, and M.-L. Newell, *HIV incidence in rural South Africa: Comparison of estimates from longitudinal surveillance and cross-sectional cBED assay testing*, PLoS ONE **3** (2008), no. 11, e3640, (Reproduced in this document as Chapter 6).

[11] _____, *Comparison of HIV incidence and risk factors for recent infection based on longitudinal follow-up versus cross-sectional cBED assay testing: a study in rural South Africa*, 16th Conference on Retroviruses and Opportunistic Infections (CROI), Montreal, Canada, 2009.

[12] T. Bärnighausen, A. Welte, T.A. McWalter, and M.-L. Newell, *Calibrating the BED capture immunoassay using data from a large population-based HIV surveillance in rural South Africa*, In preparation.

[13] Nelson KE Beyrer C, *Loss to follow-up effect in investigations of HIV-1 incidence*, Lancet **349** (1997), no. 9052, 649–50.

[14] R. Brookmeyer, *Response to correspondence on 'Should biomarker estimates of HIV incidence be adjusted?'*, AIDS **23** (2009), no. 15, 2066–2068.

[15] _____, *Should biomarker estimates of HIV incidence be adjusted?*, AIDS **23** (2009), no. 4, 485–491.

[16] _____, *On the statistical accuracy of biomarker assays for HIV incidence*, JAIDS (2010), (In press).

[17] R. Brookmeyer and T.C. Quinn, *Estimation of current human immunodeficiency virus incidence rates from a cross-sectional survey using early diagnostic tests*, American Journal of Epidemiology **141** (1995), no. 2, 166–172.

[18] K. Buchacz, J.D. Klausner, P.R. Kerndt, R.L. Shouse, I. Onorato, P.D. McElroy, J. Schwendemann, P.B. Tambe, M. Allen, F. Coye, C. Kent, M.N. Park, K. Hawkins, E. Samoff, and J.T. Brooks, *HIV incidence among men diagnosed with early syphilis in Atlanta, San Francisco, and Los Angeles, 2004 to 2005*, Journal of Acquired Immune Deficiency Syndrome **47** (2008), no. 2, 234–40.

[19] M. Bultreys, B.S. Parekh, A. Chao, T. Dobbs, A. Dushimimana, E. Karita, R.H. Byers, R. Song, and J.S. McDougal, *Use of an IgG capture BED enzyme immunoassay to investigate recent HIV-1 seroconversions among young women in Rwanda: comparison with 24-month prospective cohort data*, International Conference on AIDS, 2004.

[20] Calypte Biomedical Corporation, *Aware(TM) BED(TM) EIA HIV-1 Incidence Test (IgG-Capture HIV-EIA). Enzyme immunoassay for population estimates of HIV-1 incidence. cat no. 98003*, Portland, OR, USA, 2008.

[21] _____, *Incidence testing technologies*, Available at: `http://www.calypte.com/technologies-incidence-testing.asp`, 2009.

[22] D.D. Celentano, K.C. Bond, C.M. Lyles, S. Eiumtrakul, V.F. Go, C. Beyrer, C. na Chiangmai, K.E. Nelson, C. Khamboonruang, and C. Vaddhanaphuti, *Preventive intervention to reduce sexually transmitted infections: a field trial in the Royal Thai Army*, Archives of Internal Medicine **160** (2000), 535–540.

[23] Centers for Disease Control and Prevention (CDC), *BED-CEIA incidence and adjustment formula*, Available at: `http://www.cdc.gov/NCHSTP/OD/GAP/docs/surveillance/BEDCEIA%20Incidence%20and%20Adjustment%20Formula.pdf`, 2006.

[24] ———, *Interim recommendations for the use of the BED capture enzyme immunoassay for incidence estimation and surveillance.*, Available from: `http://www.cdc.gov/nchstp/od/GAP/docs/surveillance/Interim%20Recommendations%20for%20the%20use%20of%20the%20BED%20capture%20enzyme%20immunoassay%20for%20incidence%20estimation%20and%20surveillance%20Approved%20November%2021%202006%20(2).pdf`, 2006.

[25] ———, *Using the BED HIV-1 capture EIA assay to estimate incidence using STARHS in the context of surveillance in the United States*, Available at: `http://www.cdc.gov/hiv/topics/surveillance/resources/factsheets/pdf/bed.pdf`, 2007.

[26] S.R. Cole, H. Chu, and R. Brookmeyer, *Confidence intervals for biomarker-based human immunodeficiency virus incidence estimates and differences using prevalent data*, American Journal of Epidemiology **165** (2006), no. 1, 94–100.

[27] DHS, *All surveys by country*, Available: `http://www.measuredhs.com/aboutsurveys/search/search_survey_main.cfm?SrvyTp=country`. Accessed 10 May 2007.

[28] T. Dobbs, S. Kennedy, C.-P. Pau, J.S. McDougal, and J.S. Parekh, *Performance characteristics of the Immunoglobulin G-capture BED-enzyme immunoassay, an assay to detect recent human immunodeficiency virus type 1 seroconversion*, Journal of Clinincal Microbiology **42** (2004), no. 6, 2623–2628.

[29] Brillinger DR, *The natural variability of vital rates and associated statistics*, Biometrics **42** (1986), 693–734.

[30] P.D. Ghys, E. Kufa, M.V. George, for the UNAIDS Reference Group on Estimates Modelling, Projections ad hoc Working Group on interpreting trends in prevalence, and incidence of HIV infection in countries with generalised epidemics., *Measuring trends in prevalence and incidence of HIV infection in countries with generalised epidemics*, Sexually Transmitted Infections **82** (2006), no. Suppl 1, i52–i56.

[31] P.D. Ghys, B. Zaba, and M. Prins, *Survival and mortality of people infected with HIV in low and middle income countries: results from the extended ALPHA network*, AIDS **21** (2007), no. Suppl 6, S1–4.

[32] R.H. Gray, X. Li, G. Kigozi, D. Serwadda, H. Brahmbhatt, F. Wabwire-Mangen, F. Nalugoda, M. Kiddugavu, N. Sewankambo, T.C. Quinn, S.J. Reynolds, and M.J. Wawer, *Increased risk of incident HIV during pregnancy in Rakai, Uganda: a prospective study*, Lancet **366** (2005), 1182–1188.

[33] S. Gregson, C.A. Donnelly, C.G. Parker, and R.M. Anderson, *Demographic approaches to the estimation of incidence of HIV-1 infection among adults from age-specific prevalence data in stable endemic condition*, AIDS **10** (1996), 1689–1697.

[34] S.B. Gupta, G. Murphy, E. Koenig, C. Adon, C. Beyrer, D. Celentano, S. Khawaja, F. Sifakis, J.V. Parry, and W. Straus, *Comparison of methods to detect recent HIV type 1 infection in cross-sectionally collected specimens from a cohort of female sex workers in the Dominican Republic*, AIDS Research and Human Retroviruses **23** (2007), no. 12, 1475–80.

[35] H.I. Hall, R. Song, P. Rhodes, J. Prejean, Q. An, L.M. Lee, J. Karon, R. Brookmeyer, E.H. Kaplan, M.T. McKenna, and R.S. Janssen, *Estimation of HIV incidence in the United States*, JAMA **300** (2008), no. 5, 520–9.

[36] T.B. Hallett, P. Ghys, T. B arnighausen, P. Yan, and G.P. Garnett, *Errors in 'BED'-derived estimates of HIV incidence will vary by place, time and age*, PLos ONE **4** (2009), no. 5, e5720.

[37] T.B. Hallett, S. Gregson, E. Gonese, O. Mugurungi, and G.P. Garnett, *Assessing evidence for behaviour change affecting the course of HIV epidemics: A new mathematical modelling approach and application to data from Zimbabwe*, Epidemics (2009), doi:10.1016/j.epidem.2009.03.001.

[38] T.B. Hallett, B. Zaba, J. Todd, B. Lopman, W. Mwita, S. Biraro, S. Gregson, J.T. Boerma, and ALPHA Network, *Estimating incidence from prevalence in generalised HIV epidemics: methods and validation.*, Plos Medicine **5** (2008), no. 4, e80.

[39] J.W. Hargrove, *BED estimates of HIV incidence must be adjusted*, AIDS **23** (2009), no. 15, 2061–2062.

[40] J.W. Hargrove, J.H. Humphrey, K. Mutasa, B.S. Parekh, McDougal J.S., R. Ntozini, H. Chidawanyika, L.H. Moulton, B. Ward, K. Nathoo, P.J. Iliff, and E. Kopp, *Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay*, AIDS **22** (2008), no. 4, 511–518.

[41] T. Hayashida, H. Gatanaga, J. Tanuma, and S. Oka, *Effects of low HIV type 1 load and antiretroviral treatment on IgG-capture BED-enzyme immunoassay*, AIDS Research and Human Retroviruses **24** (2008), no. 3, 495–8.

[42] K. Herbst, G. Cooke, T. Bärnighausen, A. Kany, F. Tanser, and M.-L. Newell, *Adult mortality and antiretroviral treatment roll-out in a rural population in KwaZulu-Natal*, submitted, 2008.

[43] W. Hladik, D. Olara, W. Were, J. Mermin, and R. Downing, *The effect of antiretroviral treatment on the specificity of the serological BED HIV-1 incidence assay*, The 2007 HIV/AIDS PEPFAR Implementers' Meeting. Kigali, Rwanda, (Abstract 998), 2007.

[44] D.J. Hu, S. Vanichseni, P.A. Mock, N.L. Young, T. Dobbs, R.H. Jr. Byers, K. Choopanya, F. van Griensven, D. Kitayaporn, J.S. McDougal, J.W. Tappero, T.D. Mastro, and B.S. Parekh, *HIV type 1 incidence estimates by detection of recent infection from a cross-sectional sampling of injection drug users in Bangkok: use of the IgG capture BED enzyme immunoassay*, AIDS Research and Human Retroviruses **19** (2003), no. 9, 727–30.

[45] J.B. Hubert, M. Burgard, E. Dussaix, C. Tamalet, C. Deveau, J. Le Chenadec, M.L. Chaix, E. Marchadier, J.L. Vildé, J.F. Delfraissy, L. Meyer, and Rouzioux, *Natural history of serum HIV-1 RNA levels in 330 patients with a known date of infection. the SEROCO Study Group*, AIDS **14** (2000), no. 2, 123–31.

[46] International AIDS Society: abstract search, Available at: `http://www.iasociety.org/AbstractSearch.aspx`, 2009.

[47] R.S. Janssen, G.A. Satten, S.L. Stramer, B.D. Rawal, T.R. O'Brien, B.J. Weiblen, F.M. Hecht, N. Jack, F.R. Cleghorn, J.O. Kahn, M.A. Chesney, and M.P. Busch, *New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes*, JAMA **280** (1998), no. 1, 42–48.

[48] Y. Jiang, M. Wang, M. Ni, S. Duan, Y. Wang, J. Feng, Y. Xiao, Y. Dong, D. Wang, M. Han, L. Xiang, L. Ma, and Q. Zhou, *HIV-1 incidence estimates using IgG-capture BED-enzyme immunoassay from surveillance sites of injection drug users in three cities of China*, AIDS **21** (2007), no. Suppl 8, S47–S51.

[49] K. Kana, S. Tabprasit, P. Chaitaveep, P. Tienamporn, and T. Chuenchitra, *HIV-1 incidence estimates among Royal Thai Army conscripts: use of the IgG-capture BED-enzyme immunoassay*, XVII International AIDS Conference, Mexico City, Mexico, 2008.

[50] S.H. Kapiga, E.F. Lyamuya, G.K. Lwihula, and D.J. Hunter, *The incidence of hiv infection among women using family planning methods in Dar es Salaam, Tanzania*, AIDS **12** (1998), 75–84.

[51] E.H. Kaplan and R. Brookmeyer, *Snapshot estimators of recent HIV incidence rates*, Operations Research **47** (1999), no. 1, 29–37.

[52] E. Karita, M. Price, E. Hunter, E. Chomba, S. Allen, L. Fei, A. Kamali, E.J. Sanders, O. Anzala, M. Katende, and N. Ketter, *Investigating the utility of the HIV-1 BED capture enzyme immunoassay using cross-sectional and longitudinal seroconverter specimens from Africa*, AIDS **21** (2007), no. 4, 403–8.

[53] J.M. Karon, R. Song, R. Brookmeyer, E.H. Kaplan, and H.I. Hall, *Estimating HIV incidence in the United States from HIV/AIDS surveillance data and biomarker HIV test results*, Statistics in Medicine **27** (2008), no. 23, 4617–33.

[54] R. Kassanjee, A. Welte, T.A. McWalter, J. Viljoen, T. Bärnighausen, M.-L. Newell, and L.P. Fatti, *Calibration of BED assay for use in incidence estimation*, Fifth IAS Conference on HIV Pathogenesis, Treatment and Prevention, Cape Town, South Africa, CDB018, 2009.

[55] A. Kim, S. McDougal, J. Hargrove, M. Morgan, M. Nolan, L. Marum, A. Abdullahi, J. Humphrey, J. Humphrey, K. Mutasa, and B. Parekh, *Toward more plausible estimates of HIV-1 incidence in cross-sectional serologic surveys in Africa: application of a HIV-1 incidence assay with post-assay adjustments*, 14th Conference on Retroviruses and Opportunistic Infections (CROI), Los Angeles, California, USA, 2007.

[56] O. Laeyendecker, *Measuring new HIV infections; where we are with new technologies and approaches measuring HIV incidence. BED+ avidity testing algorithm*, The 2nd Global HIV/AIDS Surveillance Meeting, Bangkok, Thailand, 2009.

[57] O. Laeyendecker, R.E. Rothman, C. Henson, B.J. Horne, K.S. Ketlogetswe, C.K. Kraus, J. Shahan, G.D. Kelen, and T.C. Quinn, *The effect of viral suppression on cross-sectional incidence testing in the Johns Hopkins Hospital Emergency Department*, Journal of Acquired Immune Deficiency Syndrome **48** (2008), no. 2, 211–5.

[58] S.W. Li, X.Y. Zhang, X.X. Li, M.J. Wang, D.L. Li, Y.H. Ruan, X.X. Zhang, and Y.M. Shao, *Detection of recent HIV-1 infections among men who have sex with men in Beijing during 2005-2006*, Chinese Medical Journal (English) **121** (2008), no. 12, 1105–8.

[59] S. Loschen, J. Bätzing-Feigenbaum, G. Poggensee, C. Cordes, B. Hintsche, M. Rausch, S. Dupke, S. Gohlke-Micknis, J. Rödig, O. Hamouda, and C. Kücherer, *Comparison of the human immunodeficiency virus (HIV) type 1-specific immunoglobulin G capture enzyme-linked immunosorbent assay and the avidity index method for identification of recent HIV infections*, Journal of Clinical Microbiology **46** (2008), 341–345.

[60] D. Low-Beer, H. Afkhami, R. Komatsu, P. Banati, M. Sempala, I. Katz, J. Cutler, P. Schumacher, R. Tran-Ba-Huy, and B. Schwartl ander, *Making performance-based funding work for health*, PLoS Medicine **4** (2007), no. 8, e219.

[61] E.T. Marinda, J.W. Hargrove, W. Preiser, H. Slabbert, G. van Zyl, J. Levin, L.H. Moulton, A. Welte, and J. Humphrey, *Significantly diminished long-term specificity of the BED capture enzyme immunoassay among patients with HIV-1 with very low CD4 counts and those on antiretroviral therapy*, JAIDS **53** (2010), no. 4, 496–9.

[62] J.S. McDougal, *BED estimates of HIV incidence must be adjusted*, AIDS **23** (2009), no. 15, 2064–2065.

[63] J.S. McDougal, B.S. Parekh, M.L. Peterson, B.M Branson, T. Dobbs, M. Ackers, and M. Gurwith, *Comparison of HIV type 1 incidence observed during longitudinal follow-up with incidence estimated by cross-sectional analysis using the BED capture enzyme immunoassay*, AIDS Research and Human Retroviruses **22** (2006), no. 10, 945–952.

[64] J.S. McDougal, C.D. Pilcher, B.S. Parekh, Gershy-Damet G., K. Branson B.M., Marsh, and S.Z. Wiktor, *Surveillance for HIV-1 incidence using tests for recent infection in resource-constrained countries*, AIDS **19** (2005), no. Suppl 2, S25–S30.

[65] T.A. McWalter, R. Kassanjee, and A. Welte, *Incidence from cross-sectional surveys: Improved characterization of tests for recent infection*, XVIII International AIDS Conference. Vienna, Austria, 2010; (Abstract CDC0473), 2010, (Reproduced in this document as Chapter 9).

[66] T.A. McWalter, O. Laeyendecker, and A. Welte, *Evaluating performance of cross-sectional incidence estimators: A challenge for assay development*, South African AIDS conference, Durban, South Africa, 2009.

[67] T.A. McWalter and A. Welte, *On the estimation of the proportion of true recent infections using the BED capture enzyme immunoassay*, Available at `http://arxiv.org/abs/0806.1297`, (Reproduced in modified form as Chapter 3), June 2008.

[68] _____, *Relating recent infection prevalence to incidence with a sub-population of non-progressors.*, Available at `http://arxiv.org/abs/0801.3380`, (Reproduced in this document as Chapter 2), June 2008.

[69] _____, *A comparison of biomarker based incidence estimators*, PLoS ONE **4** (2009), no. 10, e7368, (Reproduced in this document as Chapter 3).

[70] _____, *A new paradigm for incidence estimation from cross-sectional data*, Fifth IAS Conference on HIV Pathogenesis, Treatment and Prevention, Cape Town, South Africa, MOPDB105, 2009, (Reproduced in this document as Appendix B).

[71] _____, *Relating recent infection prevalence to incidence with a sub-population of assay non-progressors*, Journal of Mathematical Biology **60** (2010), no. 5, 687–710, (Reproduced in this document as Chapter 2).

[72] Medical subject headings, Available at: `http://www.nlm.nih.gov/mesh/`, 2009.

[73] J. Mermin, J. Musinguzi, A. Opio, W. Kirungi, J.P. Ekwaru, W. Hladik, F. Kaharuza, R. Downing, and R. Bunnell, *Risk factors for recent HIV infection in Uganda*, JAMA **300** (2008), no. 5, 540–9.

[74] G. Murphy and J.V. Parry, *Assays for the detection of recent infections with human immunodeficiency virus type 1*, Eurosurveillance **13** (2008), no. 7-9, 4–10.

[75] National AIDS and STI Control Programme, Ministry of Health - Kenya, *Kenya aids indicator survey, 2007: Preliminary report*, KAIS presentation, 7th October, George Washington University, Washington D.C., USA, 2007.

[76] National Library of Medicine (NLM) Gateway. HIV/AIDS., Available at: `http://gateway.nlm.nih.gov/gw/Cmd?FAQ.x#AIDS`, 2009.

[77] K.E. Nelson, N. Galai, M. Safaeian, S.A. Strathdee, D.D. Celentano, and D. Vlahov, *Temporal trends in the incidence of human immunodeficiency virus infection and risk behavior among injection drug users in Baltimore, Maryland, 1988-1998*, American Journal of Epidemiology **156** (2002), 641–653.

[78] S. Nesheim, B. Parekh, K. Sullivan, M. Bulterys, T. Dobbs, M. Lindsay, M. Cashat-Cruz, B. Byers, and F. Lee, *Temporal trends in HIV type 1 incidence among inner-city childbearing women in Atlanta: use of the IgG-capture BED-enzyme immunoassay*, AIDS Research and Human Retroviruses **21** (2005), no. 6, 537–44.

[79] B.S. Parekh, S.M. Kennedy, T. Dobbs, C.-P. Pau, R. Byers, T. Green, D.J. Hu, S. Vanichseni, N.L. Young, K. Choopanya, T.D. Mastro, and J.S. McDougal, *Quantitative detection of increasing HIV type 1 antibodies after seroconversion: A simple assay for detecting recent HIV infection and estimating incidence*, AIDS Research and Human Retroviruses **18** (2002), no. 4, 295–307.

[80] B.S. Parekh and J.S. McDougal, *New approaches for detecting recent HIV-1 infection*, AIDS Review **3** (2001), 183–193.

[81] ———, *Application of laboratory methods for estimation of HIV-1 incidence*, Indian Journal of Medical Research **121** (2005), no. 4, 510–8.

[82] T. Plipat, A. Teeraratkul, W. Chalermchan, D. Areechokchai, A. Sriburi, V. Pobkeeree, J. Danyothapolchai, B. Parekh, K. Fox, S. Sombat Thanprasertsuk, J. Tappero, and K. Ungchusak, *National HIV incidence surveillance using BED capture immunoassay among pregnant women and female sex workers, Thailand*, XVI International AIDS Conference, Toronto, Canada, 2006.

[83] S.J. Posner, L. Myers, S.E. Hassig, J.C. Rice, P. Kissinger, and T.A. Farley, *Estimating HIV incidence and detection rates from surveillance data*, Epidemiology **15** (2004), no. 164-172.

[84] F.H. Priddy, C.D. Pilcher, R.H. Moore, P. Tambe, M.N. Park, S.A. Fiscus, M.B. Feinberg, and C. del Rio, *Detection of acute HIV infections in an urban HIV counseling and testing population in the United States*, Journal of Acquired Immune Deficiency Syndrome **44** (2007), no. 2, 196–202.

[85] PubMed, Available at: `http://www.ncbi.nlm.nih.gov/pubmed/`, 2009.

[86] T. Rehle, O. Shisana, V. Pillay, K. Zuma, A. Puren, and W. Parker, *South african medical journal*, National HIV incidence measures — new insights into the South African epidemic **97** (2007), no. 3, 194–9.

[87] C. Sakarovitch, F. Rouet, G. Murphy, A.K. Minga, A. Alioum, F. Dabis, D. Costagliola, R. Salamon, J.V. Parry, and F. Barin, *Do tests devised to detect recent HIV-1infection provide reliable estimates of incidence in Africa?*, Jounral of Acquired Immune Deficiency Syndrome **45** (2007), no. 1, 115–22.

[88] V. Saphonn, B.S. Parekh, T. Dobbs, C. Mean, L.H. Bun, S.P. Ly, S. Heng, and Detels R., *Trends of HIV-1 seroincidence among HIV-1 sentinel surveillance groups in Cambodia, 1999–2002*, Jounral of Acquired Immune Deficiency Syndrome **39** (2005), no. 5, 587–92.

[89] V. Saphonn, K. Shafer, J. Kaldor, C. Kaoeun, S. Mam, S. Sim, R. Olrich, E. Stein, C. Mom, and C. Mean, *Risk factors and HIV incidence among voluntary counselling and testing site clients in Phnom Penh, Cambodia: 2006*, 4th IAS Conference on HIV Pathogenesis, Treatment and Prevention, Sydney, Australia, 2007.

[90] L.A. Shafer, S. Biraro, J. Nakiyingi-Miiro, A. Kamali, Ssematimba D., J. Ouma, A. Ojwiya, P. Hughes, L. Van der Paal, J. Whitworth, A. Opio, and H. Grosskurth, *HIV prevalence and incidence are no longer falling in southwest Uganda: evidence from a rural population cohort 1989-2005*, AIDS **22** (2008), 1641–1649.

[91] O. Shisana, T. Rehle, L.C. Simbayi, W. Parker, S. Jooste, V. Pillay-van Wyk, N. Mbelle, and J. van Zyle, *South African national HIV prevalence, HIV incidence, behaviour and communication survey, 2008: A turning tide among teenagers?*, HSRC Press (Cape Town), 2009.

[92] O. Shisana, T. Rehle, L.C. Simbayi, W. Parker, K. Zuma, A. Bhana, C. Connolly, S. Jooste, and V. Pillay, *South African national HIV prevalence, HIV incidence, behaviour and communication survey*, HSRC Press (Cape Town), 2005.

[93] L.C. Simbayi, T. Rehle, J. Vass, D. Skinner, K. Zuma, M.N. Mbelle, S. Jooste, V. Pillay, N. Dwadwa-Henda, Y. Toefy, P. Dana, T. Ketye, and A. Matevha, *The impact of and response to HIV/AIDS in the private security and legal services industries in South Africa*, Cape Town, South Africa, 2007.

[94] C. Sinthuwattaniwibool, P. Wasinrapee, T. Chaowanachan, A. Sriinsut, B. Parekh, S. McDougal, J. Tappero, and J. McNicholl, *Application of the BED-CEIA to a cohort of military conscripts in northern Thailand to estimate HIV-1 incidence: a validation study*, 12th Conference on Retroviruses and Opportunistic Infections (CROI), Boston, Massachusetts, USA, 2005.

[95] South African National Strategic plan for HIV AIDS, Available from: `http://www.sanac.org.za/National%20Strategic%20Plan%20on%20HIV,%20AIDS%20&%20STI%202007%20-%202011.pdf`, 2007.

[96] A.K. Srikrishnan, S.S. Solomon, C.K. Vasudevan, E. Thamburaj, C. Latkin, M. Suresh kumar, K. Murugavel, H. Syed Iqbal, and D. Celentano, *Incidence and prevalence of HIV and coinfections among injection drug users in Chennai, India — cause for concern*, XVI International AIDS Conference, Toronto, Canada, 2006.

[97] J. Stover, P. Johnson, B. Zaba, M. Zwahlen, F. Dabis, and Ekpini R.E., *The spectrum projection package: improvements in estimating mortality, art needs, pmtct impact and uncertainty bounds.*, Sexually Transmitted Infections **84** (2008), no. Suppl 1, i24–i30.

[98] B. Suligoi, C. Galli, M. Massi, F. Di Sora, M. Sciandra, P. Pezzotti, O. Recchia, F. Montella, A. Sinicco, and G. Rezza, *Precision and accuracy of a procedure for detecting recent human immunodeficiency virus infections by calculating the antibody avidity index by an automated immunoassay-based method*, Journal of Clinical Microbiology **40** (2002), no. 11, 4015–20.

[99] F. Tanser, V. Hosegood, T. Bärnighausen, K. Herbst, M. Nyirenda, W. Muhwava, C. Newell, J. Viljoen, and T. Newell M.-L. Mutevedzi, *Cohort profile: Africa centre demographic information system (ACDIS) and population-based HIV survey*, International Journal of Epidemiology **37** (2007), 956–962.

[100] J. Todd, J.R. Glynn, M. Marston, T. Lutalo, S. Biraro, W. Mwita, V. Suriyanon, R. Rangsin, K.E. Nelson, P. Sonnenberg, D. Fitzgerald, E. Karita, and B. Zaba, *Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy*, AIDS **21** (2007), no. Suppl 6, S55–S63.

[101] H.-H.M. Truong, K. Fritz, W. McFarland, W. Hartogenesis, A. Fiamma, T. Coates, and S. Morin, *Recent HIV infections detected at a mobile HIV voluntary counseling and testing program in Zimbabwe*, 14th Conference on Retroviruses and Opportunistic Infections (CROI), Los Angeles, California, USA, 2007.

[102] H.-H.M. Truong, T. Kellogg, B. Louie, J. Klausner, J. Dilley, and W. McFarland, *Comparison of HIV incidence estimates derived from laboratory incidence assays and repeat testing data at HIV testing sites in San Francisco*, 14th Conference on Retroviruses and Opportunistic Infections (CROI), Los Angeles, California, USA, 2007.

[103] H.-H.M. Truong, M. Mathur, S. Mungekar, J. Moncada, A. Hernandez, H. Jerajani, and C. Lindan, *High rates of recent HIV-1 infections among male attendees of public STI clinics in Mumbai are associated with visits to commercial sex workers*, XVII International AIDS Conference, Mexico City, Mexico, 2008.

[104] H.-H.M. Truong, P. Samnang, K. Page-Shafer, B. Louie, W. McFarland, and R.M. Grant, *Concordance of STARHS and BED assays for identifying recent HIV infections in Cambodian fishermen*, XV International AIDS Conference, Bangkok, Thailand, 2004.

[105] UNAIDS Reference Group on Estimates Modelling and Projections, *Statement on the use of the BED-assay for the estimation of HIV-1 incidence for surveillance or epidemic monitoring*, Available at: `http://data.unaids.org/pub/EPISlides/2006/Statement_BED_Policy_13Dec05_en.pdf`, 2005.

[106] A van den Bos, *Parameter estimation for scientists and engineers.*, John Wiley & Sons, 2007.

[107] C.A. Velasco de Castro, B. Grinsztejn, V. Veloso, F.I. Bastos, J.H. Pilotto, R. Friedman, R.I. Moreira, and M.G. Morgado, *Estimated HIV incidence and diversity among men who have sex with men (MSM) in VCTs of Rio de Janeiro State, Brazil*, 4th IAS Conference on HIV Pathogenesis, Treatment and Prevention, Sydney, Australia, 2007.

[108] C.A. Velasco de Castro, B. Grinsztejn, V. Veloso, F.I. Bastos, J.H. Pilotto, M. Paiva, R. Friedman, R. Moreira, and M.G. Morgado, *HIV-1 incidence, prevalence and socialdemographic profile in VCTs of Rio de Janeiro State, Brazil*, XVI International AIDS Conference, Toronto, Canada, 2006.

[109] S. Verevochkin, A. Masharsky, A. Shaboltas, R. Heimer, I. Hoffman, and A. Kozlov, *Subtype A HIV-1 infection incidence estimates: differences between lab assays for detection of recent HIV infections and correlation with cohort studies among IDU in St.Petersburg, Russia*, XVII International AIDS Conference, Mexico City, Mexico, 2008.

[110] N. Wang and J. Xu, *Trends in HIV-incidence rates by the BED-CEIA/HIV pooled-PCR in a longitudinal cohort survey of FSWs of Kaiyuan city, China*, XVII International AIDS Conference, Mexico City, Mexico, 2008.

[111] N. Wang, J.J. Xu, L. Lu, J.G. Ma, H.B. Wang, W. Michelle, Z.L. Wu, and X.W. Zheng, *Application of BED-CEIA and prospective cohort study for HIV incidence estimation among FSWs in Yunnan, China*, 4th IAS Conference on HIV Pathogenesis, Treatment and Prevention, Sydney, Australia, 2007.

[112] R. Wang and S.W. Lagakos, *On the use of adjusted cross-sectional estimators of HIV incidence*, JAIDS **52** (2009), no. 5, 538–547.

[113] P. Wasinrapee, T. Chaowanachan, M. Martin, T. Dobbs, B. Parekh, L. Srisuwanwilai, K. Choopanya, S. Vanichseni, D. Hu, F. van Griensven, A. Luck, and J. Tappero, *Estimation of HIV-1 incidence from a cross-sectional assessment of injecting drug users using the HIV-1 subtype BED IgG-capture EIA*, XV International AIDS Conference, Bangkok, Thailand, 2004.

[114] P. Wasinrapee, J. McNicholl, B. Parekh, and S. McDougal, *Validation of HIV-1 incidence estimation with the immunoglobulin G-capture BED-enzyme immunoassay (BED assay) in three longitudinal cohorts from Thailand*, XVI International AIDS Conference, Toronto, Canada, 2006.

[115] A. Welte, *Relating incidence to 'recent infection' prevalence: Application to HIV*, South African Journal of Science **5-6** (2008), no. 104, 199–202.

[116] A. Welte, T.A. McWalter, and T. Bärnighausen, *Reply to 'Should biomarker estimates of HIV incidence be adjusted?'*, AIDS **23** (2009), no. 15, 2062–2063, (Reproduced in this document as Chapter 8).

[117] ———, *A simplified formula for inferring HIV incidence from cross-sectional surveys using a test for recent infection*, AIDS Research and Human Retroviruses **25** (2009), no. 1, 125–126, (Reproduced in this document as Chapter 4).

[118] A. Welte, T.A. McWalter, O. Laeyendecker, and T.B. Hallet, *Using tests for recent infection to estimate incidence: problems and prospects for HIV.*, Eurosurveillance **15** (2010), no. 24, pii=19589, (Reproduced in this document as Chapter 5).

[119] T. Welz, V. Hosegood, S. Jaffar, J. B atzing-Feigenbaum, K. Herbst, and M.-L. Newell, *Continued very high prevalence of HIV infection in rural KwaZulu-Natal, South Africa: a population-based longitudinal study*, AIDS **21** (2007), 139–144.

[120] B. Williams, E. Gouws, D. Wilkinson, and S.A. Karim, *Estimating HIV incidence rates from age prevalence data in epidemic situations*, Statistics in Medicine **20** (2001), 2003–2016.

[121] D. Wolday, H Meles, E. Hailu, T. Messele, Y. Mengistu, M. Fekadu, B.S. Parekh, and T. Wuhib, *Temporal trends in the incidence of HIV infection in antenatal clinic attendees in Addis Ababa, Ethiopia, 1995–2003*, Journal of Internal Medicine **261** (2007), no. 2, 132–137.

[122] K.-F. Wong, W.-Y. Tsai, and L. Kuhn, *Estimating HIV hazard rates from cross-sectional HIV prevalence data*, Statistics in Medicine **25** (2006), no. 15, 2441–2449.

[123] World Health Organisation, *WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children*, Geneva: WHO, 2007.

[124] Y. Xiao, Y. Jiang, J. Feng, W. Xu, M. Wang, E. Funkhouser, S.H. Vermund, and Y. Jia, *Seroincidence of recent human immunodeficiency virus type 1 infections in China*, Clinical Vaccine Immunology **14** (2007), no. 10, 1384–6.