Overview of Incidence Estimation Package *inctools*

29 April 2017

Introduction

The *inctools* package is broadly conceived to provide state of the art functionality to support numerous aspects of population level incidence surveillance. Estimating incidence in formally constituted study cohorts is conceptually straightforward and not the focus of this package.

Inspiration for this work derives from the challenges associated with estimating population level HIV incidence, and so, for the time being, this will explicitly be the primary focus and the sole source of examples. In principle, the estimation of any non-remissible condition poses similar obstacles, so it is plausible that these ideas may find fruitful application to conditions other than HIV.

In this initial version of *inctools* , the central theme is the estimation of incidence through cross-sectional biomarker measurement – in particular, biomarkers supporting classification of cases (infections) as recently or non-recently acquired. In future, it is expected that there will be functionality to explicitly address the interpretation of age and time structured prevalence and mortality data, whether or not 'recency of infection' has been ascertained.

This package will be updated from time to time, with releases spun off as branches from a git repository housed at (https://github.com/SACEMA/inctools). The full repository is freely accessible under GNU General Public Licence (version 3). In other words, anyone may freely use, distribute or modify this material as long as the original creators/contributors are appropriately acknowledged and all derivative work released is covered by the same licence. This does not extend to computational outputs from these tools, over which the tool creators make no claim.

Potential users include: * researchers planning, implementing, or analysing data from, major surveys * officers in departments of health or statistical bureaux * reviewers of protocols or articles * product developers * funders * teachers/trainers and students

While the correct use of *inctools* does not require any detailed technical knowledge about the theoretical underpinnings, a solid grasp of the conceptual landscape is probably necessary. This vignette should help potential users orient themselves and identify gaps for further background exploration. The other vignettes demonstrate the use of the key functions. To support correct use of the tool, there are numerous warnings and error messages designed to appear in response to a variety of triggers.

- **Warnings** highlight possible problems such as input parameters that are unlikely to be consistent with robust applications.
- **Errors** indicate a fundamental breakdown in consistency, such as a greater number of recent HIV infections than HIV infections. In the case of an error, expected output is suppressed.

The lack of any formal warranty notwithstanding, users are invited to contact the maintainers for assistance and are indeed requested to provide any notification of possible errors, feedback about functionality or notification and explanation of GIT repository level 'pull-requests'.

Historical context

Amongst other features, Version 1 of *inctools* makes available the functionality of a spreadsheet based tool set previous circulated in three major versions under the name ABIE (*Assay Based Incidence Estimation*). Version 3 of ABIE is still available and may for a little while retain some advantages over an R package, for some users, such as the automatic generation of plots as part of some calculations, the production of which, in the initial R package, would require some R programmer intervention. However, the ABIE spreadsheet tools family is now technically obsolete and superseded, but will be supported for a while on [www.incidence-estimation.org]

(http://www.incidence-estimation.org/page/tools-for-incidence-from-biomarkers-for-recent-infection). The ABIE feature set forms the basis of two components of *inctools*: [^1]: Analysis in support of survey design [^2]: Analysis of survey data Additionally, *inctools* incorporates the heart of a little known, now deprecated, R package called ritcalib, which provides [^3]: Estimation of key performance characteristics of tests for 'recent infection'

Overview of Functionality and Use

Fleshing out the three primary components just noted above:

Analysis in support of survey design

Two primary exposed functions (`incprecision` and `incpower`) provide permutations on the relationship between

- epidemiological context,
- sampling frame,
- test properties, and
- statistical informativeness

(expressed through incidence estimate *precision*, or trend/difference detection *power*, respectively). The primary (and currently implemented) uses are expected to be

- the calculation of either precision or power from the other contextual factors
- the calculation of a sample size from the other contextual factors, including desired precision or power

Users of `incprecision` and `incpower` will need to make significant efforts to assess the context of intended use, but require no detailed knowledge of the statistical theory underpinning the details of the calculations. The two primary functions each offer their various respective permutations on usage by having the user provide the value "out" for the parameter which is requested as output, while all the other logically necessary parameters will need to be supplied, in order to avoid an error.

Analysis of survey data

Once a population level survey has been carried out, the fully specified data set will essentially consist of subject level records indicating demographic factors, cluster/stratum membership, and various clinical indicators such as final HIV status and recency status. In order to accommodate arbitrarily sophisticated analysis of such complex data sets, the conception of *inctools* is that a substantial layer of 'pre-processing' is expected to routinely precede any invocation of *inctools* functionality. Use of the primary, preferred entry point into incidence estimate (function incprops) requires survey data to be reduced to:

- an estimate of HIV prevalence
- an estimate of the prevalence of 'recent infection' amongst those who are HIV positive.
- A variance/covariance matrix for these two prevalence estimates

While *inctools* makes no firm statements about how these estimates are to be obtained, it is recommended that they be provided in the specified form, and that function `incprops` be used as the primary entrypoint. Note that:

- To facilitate "naïve" analysis, the function `prevcounts` is provided, which turns simple sample case counts (number of people tested for HIV, number classified HIV positive, number of HIV positives tested for recency, number classified as recently infected) into the prevalence estimates required by function incprops, under the assumptions of independent (unstructured) sampling.

- Function inccounts incorporates a detour through function prevcounts, followed by a call to incprops, thereby provide a high level interface to incidence estimation based directly on total sample counts, contextual factors, and recency test properties.
- The package *inctools* lists package survey in its dependencies, thereby assuring access to at least one robust flexible tool set to facilitate the estimation of the relevant sample proportions from structured survey data.

Estimation of key performance characteristics of *Tests for Recent Infection*

It is usually the responsibility of the developers of recent infection tests to decide on laboratory procedures and to estimate the MDRIs and (population group level) FRRs of emerging assays. Most users of such tests who are conducting field work will access these estimates from the literature. However, the process of test development is ongoing, developers at key agencies (such as NIH and CDC) are also involved in survey design and execution, and the methods applicable to survey execution are very tightly aligned to the analysis of test properties. Thus, both developers and users of recent infection tests (as well as of the analytical methodology supporting their use) are essentially part of a single broad community engaged in large scale surveillance. Therefore, it makes sense to co-locate the developers' tools with the surveillance tools.

The core functionality in this section of the package is provided by function `mdrical`, which produces estimates of mean duration of recent infection and function `frrcal` which produces estimates of false recent rate, both from data typically only available to test developers i.e. patient level test results from a putative recent infection test, applied at various times relative to relatively precisely estimated infection/seroconversion times.

Glossary

The package inctools, and natural discussion of its application, implies use of some common or specialised (to this sub-field) terms which it may be prudent to define here. These are not detailed technical definitions, but are intended to serve as reminders, or, should they be unclear, would serve to highlight the need to investigate primary sources.

HIV infected individual

The protocol-specific case definition for HIV infected individual needs to be very clearly spelled out. This is because there is no universal standard, although there have for some years been variants on protocols involving sensitive screening tests and highly specific "confirmatory" or "supplemental" tests. What could be called shifting sands in clinical practice is even more fluid in research settings, which may use viral nucleic acid and antigen detection, and not even rely on classical serological 'confirmation' of HIV infection.

Mean duration of recent infection (MDRI)

The average time for which subjects satisfy a particular "recent infection" case definition, within a specified recency cut-off time T after (detectable) infection (which is context/protocol specific).

False-recent rate (FRR)

The (context specific) fraction of tests, performed on individuals (detectably) infected for more than the time cut-off T, which produce a (false) recent result. This term has seen many variants. FRR is inspired by the long used term "error rate" to refer to the fraction of tests which fail in some sense. Note that there is fundamentally no such thing as a false non-recent result – the phenomenon that some individuals transition to the non-recent case definition at relatively early times post-infection, compared to the average time, is a

natural and normal example of inter-subject variability in response, and is accounted for in the definition of MDRI.

Incidence as a (or, an "instantaneous") rate

This is the most fundamental metric for expressing the rate at which HIV infections occur in the susceptible (aka "at risk") population, and is naturally expressed as a number of (infection) events per person time at risk in the referenced susceptible population. In the case of some other epidemiological contexts (such as influenza) it is not uncommon to refer to person time in the entire population, rather than the susceptible sub-population, although this does occasionally cause confusion. While, in principle, any unit of time may be used (days, weeks, months) the usual unit in HIV discourse is the year. It is important to note that the value of an incidence rate can in principle take any value, as it changes with choice of units in which time is measured. For contrast, see the next item immediately below.

Annual(ised) risk of infection

It is also common to report the "cumulative" probability of infection over a specific period of time, such as one year. It is a subtle point, not worth expositing here in detail, that the instantaneous incidence, with time measured in years, is not, in principle, the same value as the annual risk of infection. Suffice to repeat what was noted in the preceding glossary item, namely that incidence as a "rate" can in principle attain any value, depending on choice of units of time measurement, and then to add that, by contrast, the probability of infection cumulated over a particular time period is always a number between zero and one.

Recency time cut-off T ("Big T")

On account of the fact that it is possible for a recency test to classify some individuals as recently infected at long times post infection, the use of a time cut-off T has been introduced to assist in the housekeeping. The details of how this works are well beyond the scope of this user guide, but if using a recent infection test based on a reasonable body of literature, there is little risk to adopting previously proposed cut off time.

Relative standard error (RSE)

This widely used term refers to the ratio of a standard error of an estimate to the point estimate.

Null hypothesis

A usually artificial assumption (not necessarily strongly believed, and perhaps strongly suspected to be false) which data either falsifies or fails to falsify.

p-value

The probability, calculated under a particular null hypothesis, of seeing (at least) a specified deviation from a null value in a test-statistic under consideration. The classic p-value in this context answers the question: If the incidence were really the same in two populations which have been surveyed, what is the probability of seeing a point estimate of the incidence difference, the absolute value of which is at least as large as the one observed?

Significance

This widely used term refers to a threshold on a p-value, below which the experimenters will reject a given null hypothesis and declare there to be a "statistically significant" "effect" or "Difference".

Sample size

The total number of individuals whose HIV status has been, or is proposed to be, assessed.

Design effect

This parameter captures the impact of complexities in sampling which arise from the fact that individuals surveyed are not independently selected from the total population nominally 'represented'. These complexities include hierarchical (clustered) sampling, stratification, weighting, etc. A Design effect captures the ratio of the actual variance of a metric (such as a prevalence of some defined condition) to the variance that would be obtained for the same metric if the analysis were conducted on the simplifying assumption that the individuals surveyed have been drawn independently from a large population.

There is provision for users to provide scale factors adjusting the uncertainty in estimates of

1. the proportion of HIV infected individuals (i.e. prevalence) and
2. the proportion of "recent" results among HIV positive individuals tested for recent infection.

It should be noted that:

- There is no clear consensus on approach, and there are no mature tools, for estimating the design effect parameters for population based HIV surveys.
- The two design effects are logically independent parameters, which capture the effects of statistically independent processes.
- This package does not provide any functionality to derive or justify design effect estimates.
- It is not appropriate to scale the variances of the test property (MDRI and FRR) estimates, as these are not estimated in the incidence survey, but rather arise in the recency test development process.

Citation

The use of package *inctools* implies the use of the methodology derived in the following article, which: Kassanjee R, McWalter TA, Bärnighausen T, Welte A. A new general biomarker-based incidence estimator. Epidemiology. 2012; 23(5): 721-728. Some features implemented in *inctools* are currently not published outside the package contents, and hence some additional technical details are documented in the appendix of this vignette.

Acknowledgements

Appendix: Technical details

The critical core of all the functionality of package *inctools* is the incidence estimator of *Kassanjee et al* and some hitherto unpublished generalisations to accomodate structured survey design, including incomplete

coverage of recency ascertainment. In essence, this generalisation involves the replacement of simple survey counts, assumed in *Kassanjee et al* to be trinomially distributed, by estimated prevalences of HIV and of recent infection amongst HIV positives (including covariance, which does not arise in the trinomial distribution case). Additionally, the variance of the difference of incidence estimators raises further issues such as

- the full specification of the Null Hypothesis of equal incidence that is required for the calculatino of p values, and
- the sharing of test property estimates between surveys, which induces some correlation between the incidence estimators calculated from those surveys.

Three cases of test property sharing can feasibly arise:

1. A single estimate of MDRI, and a single estimate of FRR, is available for the recency tests used in multiple surveys. This is probably the result of the same test being used, *and* there being no reason to distinguish its FRR between the two contexts.
2. A single estimate of MDRI is shared between two surveys, but the FRRs are independently estimated. This is probably the result of the same test being used, *but* there being data to distinguish its FRR between the two contexts.
3. Both MDRI and FRR are independently estimated in each survey.

Below are the variances of the incidence difference estimator for the three cases

Case 1: Surveys share a single estimate for MDRI and a single estimate for FRR

$$V[\Delta\mathbb{I}]_1 \approx \left(\frac{(PrevR_1 - \beta)}{(1 - PrevH_1)^2}(\Omega - \beta T) \right)^2 (RSE_{PrevH_1} PrevH_1)^2 + \tag{1}$$

$$\left(\frac{(PrevR_2 - \beta)}{(1 - PrevH_2)^2}(\Omega - \beta T) \right)^2 (RSE_{PrevH_2} PrevH_2)^2 + \tag{2}$$

$$\frac{PrevH_1}{1 - PrevH_1}(\Omega - \beta T)) (RSE_{PrevR_1} PrevR_1)^2 + \tag{3}$$

$$\frac{PrevH_2}{1 - PrevH_2}(\Omega - \beta T)) (RSE_{PrevR_2} PrevR_2)^2 + \tag{4}$$

$$\left(\left(\frac{\beta PrevH_1 - PrevR_1 PrevH_1}{1 - PrevH_1} - \frac{\beta PrevH_2 - PrevR_2 PrevH_2}{1 - PrevH_2} \right)(\Omega - \beta T)^2 \right)^2 RSE_{MDRI} MDRI + \tag{5}$$

$$\left(\left(\frac{PrevH_1(T PrevR_1 - \Omega)}{(1 - PrevH_1)} - \frac{PrevH_2(T PrevR_2 - \Omega)}{(1 - PrevH_2)} \right)(\Omega - \beta T)^2 \right)^2 RSE_\beta \beta \tag{6}$$

$$\tag{7}$$

Case 2: Surveys share a single estimate for MDRI, but have independent FRR estimates

6

$$V[\Delta\mathbb{I}]_2 \approx \left(\frac{(PrevR_1 - \beta)}{(1 - PrevH_1)^2}(\Omega - \beta T)\right)^2 (RSE_{PrevH_1} PrevH_1)^2 + \tag{8}$$

$$\left(\frac{(PrevR_2 - \beta)}{(1 - PrevH_2)^2}(\Omega - \beta T)\right)^2 (RSE_{PrevH_2} PrevH_2)^2 + \tag{9}$$

$$\frac{PrevH_1}{1 - PrevH_1}(\Omega - \beta T)) (RSE_{PrevR_1} PrevR_1)^2) + \tag{10}$$

$$\frac{PrevH_2}{1 - PrevH_2}(\Omega - \beta T))(RSE_{PrevR_2} PrevR_2)^2) + \tag{11}$$

$$\left(\left(\frac{\beta\ PrevH_1 - PrevR_1\ PrevH_1}{1 - PrevH_1} - \frac{\beta\ PrevH_2 - PrevR_2\ PrevH_2}{1 - PrevH_2}\right)(\Omega - \beta\ T)^2\right)^2 RSE_{MDRI}\ MDRI + \tag{12}$$

$$\left(\frac{PrevH_1(T\ PrevR_1 - \Omega)}{1 - PrevH_1}(\Omega - \beta_1\ T)^2\right)^2 RSE_{\beta_1}\ \beta_1 + \tag{13}$$

$$\left(\frac{PrevH_2(T\ PrevR_2 - \Omega)}{1 - PrevH_2}(\Omega - \beta_2\ T)^2\right)^2 RSE_{\beta_2}\ \beta_2 \tag{14}$$

Case 3: Both MDRI and FRR are independently estimated in each survey

$$V[\Delta\mathbb{I}]_3 \approx \left(\frac{(PrevR_1 - \beta_1)}{(1 - PrevH_1)^2}(\Omega_1 - \beta_1 T_1)\right)^2 (RSE_{PrevH_1} PrevH_1)^2 + \tag{15}$$

$$\left(\frac{(PrevR_2 - \beta_2)}{(1 - PrevH_2)^2}(\Omega_2 - \beta_2 T_2)\right)^2 (RSE_{PrevH_2} PrevH_2)^2 + \tag{16}$$

$$\frac{PrevH_1}{1 - PrevH_1}(\Omega_1 - \beta_1 T_1)) (RSE_{PrevR_1} PrevR_1)^2) + \tag{17}$$

$$\frac{PrevH_2}{1 - PrevH_2}(\Omega_2 - \beta_2 T_2))(RSE_{PrevR_2} PrevR_2)^2) + \tag{18}$$

$$\left(\frac{\beta_1\ PrevH_1 - PrevR_1\ PrevH_1}{1 - PrevH_1}\ (\Omega_1 - \beta_1\ T_1)^2\right)^2 RSE_{MDRI_1}\ MDRI_1 + \tag{19}$$

$$\left(\frac{\beta_2\ PrevH_2 - PrevR_2\ PrevH_2}{1 - PrevH_2}\ (\Omega_2 - \beta_2\ T_2)^2\right)^2 RSE_{MDRI_2}\ MDRI_2 + \tag{20}$$

$$\left(\frac{PrevH_1(T_1\ PrevR_1 - \Omega_1)}{1 - PrevH_1}(\Omega_1 - \beta_1\ T_1)^2\right)^2 RSE_{\beta_1}\ \beta_1 + \tag{21}$$

$$\left(\frac{PrevH_2(T_2\ PrevR_2 - \Omega_2)}{1 - PrevH_2}(\Omega_2 - \beta_2\ T_2)^2\right)^2 RSE_{\beta_2}\ \beta_2 \tag{22}$$

Note:

- The option for the assay characteristic scheme is specified by function parameter `BMest`. Subscripts 1 and 2 in the parameter notation denote associated survey or test parameters. The variances of the three cases follow.
- To avoid unnecessarily confusing complexity in the specification of multiple *recency-test-property-estimate-sharing* rules for different subsets of a global set of surveys, only a single *recency-test-property-estimate-sharing* rule is allowed in any set of surveys submitted in one invocation of the functions `incprops` and `inccounts`.
- If different recency tests are used in separate surveys, but they happen to have any numerically identical test property estimates, this will still need to be specified as independent test properties, as these estimates would be *statistically independent*.

Crucial note on choice of null hypothesis, calculation of p-values and meaning of "statistical power"

It is not uncommon for a null hypothesis to be underspecified until the final data is available, i.e. for certain details (such as a hypothetically shared value of a test statistic) to be specified (estimated) in light of the data (such as by a "pooled estimate"). In the present situation, the hypothetically common value of incidence, as well as the (potentially different by context) values of prevalence, cannot be reasonably supplied without seeing the data. Once data is available, there is further no unique, obviously correct, approach to deriving these parameters, and different choices can be entertained. The simplest, but not necessarily optimal, (indeed potentially problematic) approach is to hypothesize that the two survey populations are identical in terms of incidence AND prevalence, as this suggests a naïve pooling of all data to provide the hypothetically common values. However, a simple test for difference of prevalence would be expected to lead to the rejection of the null hypothesis (of equal prevalence) in many, if not most, cases, as prevalence will be relatively precisely estimatable in any survey remotely powered to estimate incidence. Hence, a more flexible approach to the null hypothesis is required. The approach taken here is to:

- hypothesize that the incidence is the same in the two contexts
- be agnostic on the matter of the value of this shared incidence
- hypothesize that the estimated value of the variance of the incidence difference estimate (calculated carefully in accordance with the choice of how test properties are shared between the surveys) is a good estimate of the actual variance of the incidence difference. (This is equivalent to a routinely made assumption in less complex situations)
- allow the prevalence for each context to be independently estimated from the data from that context. (Note that even if the data does not support the rejection of the hypothesis that the prevalences are identical, benchmarking indicates that there is little or no advantage in constructing the null hypothesis to include equal prevalence.)

The preceding assumptions allow the calculation of a p-value, which is done according to a two-tailed test, capturing the point that the experimenter has no basis to pre-determine the acceptable direction of any incidence difference. Note that while it is common to report "power" as simply the probability of correctly rejecting a null hypothesis, without regard to whether the inferred direction of difference is in correspondence with the underlying reality, the present construction specifically presumes a two-tailed test, where the direction of any particular inferred difference will be either correct or incorrect. Hence, *inctools* calculates the probability of *correctly* detecting the sign of the incidence difference through the rejection of the null hypothesis throug the attainment of a below-threshold *p value*. Numerically, the probability of rejecting the null hypothesis, but with the incorrect direction for the inferred effect, will usually be *very* small, unless the study design has no useful statistical power in any case. The primary difference between this implementation and an alternative likely to be proposed, is in the insistence on a two tailed test, which should not be controversial in this situation as it would be difficult to defend the claim that survey implementers can claim to know, in advance, the direction of a change in incidence.