Estimating incidence from cross-sectional data

29 April 2017

Introduction

This vignette covers the use of the functions `incprops`, `inccounts`, and `prevcounts`. It is strongly recommended that the vignette "Introduction" be read before any use of this package.

The two primary functions for HIV incidence estimation are `incprops` and `inccounts`. These take as arguments a summary of arbitrarily complex survey data sets capturing HIV prevalence and prevalence of recent HIV infection among HIV positive subjects. They return estimates of incidence, and, if specification of multiple cross-sectional surveys is provided, incidence differences (point estimates, confidence intervals, p values, and subsidiary output). The principal reference for the methodology underlying this implementation is Kassanjee et al. *Epidemiology*, 2012.[1] Further guidance is provided in Kassanjee, McWalter, Welte. *AIDS Research and Human Retroviruses*, 2014.[2], and some hitherto unpublished technical details are in the appendix of vignette "Introduction".

Analytical Paradigm

A fundamental element in the conception of the *inctools* is that the primary entry point into the critical methodology which *inctools* implements is the function `incprops`, which takes, as summary of the population state, estimates of HIV prevalence and the prevalence of *recent infection* amongst HIV positive subjects (including variance and covariance). These estimates, in turn, would usually be best derived by (potentially complex) preliminary analysis of the raw survey data documenting individual subjects' status ascertainment, cluster and strata membership, weighting, etc.

The derivation of these prevalence estimates from raw data is in principle facilitated by various algorithms which are implemented in other packages, and are essentially independent of any of the innovation captured in this package. Use of *inctools* does not imply any specific approach to the preliminary analytical methodology, but the widely used package *survey* (totally independently maintained, with no link to *inctools*) may be suitable for many typical data sets. Additionally, to facilitate 'naive', self contained within the package, analysis, the ancilliary function `prevcounts` is provided. This takes survey counts and produces prevalence estimates (for both HIV and recent infection, including variance) under the simplifying assumption of individual level random selection of subjects from a single population group.

Using functions `incprops` and `inccounts`

The functions `incprops` and `inccounts` provide a near-identical interface, as further detailed in the help pages. Both functions take considerably pre-processed data specifying a recent infection test and a survey in which it is used: * estimates of false recency rate (FRR–β) and mean duration of recent infection (MDRI–Ω_T) and their respective relative standard errors and recency time cutoff (T) * and survey data: proportions (counts, if using function `inccounts`) of HIV positives (PrevH) and positives for recency (PrevR) and their relative standard errors.

A critical distinction is that with the use of `incprops`, variance of prevalences, including covariance, is explicitly supplied, and with the use of `inccounts`, variance emerges from counts and design effects, and there is no covariance.

---

[1]Kassanjee, R., McWalter, T.A., Baernighausen, T. and Welte, A. "A new general biomarker-based incidence estimator." Epidemiology; 2012, 23(5): 721-728.

[2]Kassanjee, R., McWalter, T.A. and Welte, A. "Short Communication: Defining Optimality of a Test for Recent Infection for HIV Incidence Surveillance." AIDS Research and Human Retroviruses; 2014, 30(1): 45-49.

The output for a single survey is an estimate of incidence along with confidence intervals and RSE, estimated annual rate of infection and associated confidence intervals, and confidence intervals for parameters MDRI and FRR, which are deduced from input parameters.

The output for multiple surveys is the same output as for a single survey, along with pairwise comparisons of incidence rates, confidence intervals of differences, and tests of equality with p-values and RSE of differences.

Examples

Consider a single cross sectional survey summarised by:

- An HIV prevalence of 20% estimated with a relative standard error of 2.8%
- An estimated prevalence of recent infection, amongst HIV positives, of 10%, with a relative standard error of 9%
- an MDRI of 200 days estimated with a relative standard error of 5%
- an FRR of 5% estimated with a relative standard error of 20%

and proposed to be processed by 10,000 bootstap iterations. Function `incprops` will calculate:

- an incidence point estimate, confidence interval, and relative standard error
- the equivalent annualised risk of infection (which is usually numerically very similar)
- implied confidence intervals on MDRi and FRR derived from input parameters

```
incprops(PrevH = 0.20, RSE_PrevH = 0.028, PrevR = 0.10, RSE_PrevR = 0.09,
        BS_Count = 10000, Boot = TRUE, MDRI = 200, RSE_MDRI = 0.05,
        FRR = 0.01, RSE_FRR = 0.2, BigT = 730)
```

```
## $Incidence.Statistics
##    Incidence CI.low   CI.up      RSE
## 1    0.04265 0.0332 0.05317 0.11948
##
## $Annual.Risk.of.Infection
##       ARI ARI.CI.low ARI.CI.up
## 1 0.0418     0.0327    0.0518
##
## $MDRI.CI
##   CI.low CI.up
## 1  180.4 219.6
##
## $FRR.CI
##   CI.low  CI.up
## 1 0.0061 0.0139
```

Multiple surveys can be processed in a single call to `incprops` by supplying vectors of the parameters. Note that:

- A single value of recency cut off time T is required for all surveys, even if the tests are different and have independently estimates MDRI and FRR, as in the example below
- In addition to the results returned for the single survey case (replicated for each survey), `incprops` also calculates all pairwise incidence differences and p values.

```
incprops(PrevH = c(0.20,0.21,0.18), RSE_PrevH = c(0.028,0.03,0.022),
        PrevR = c(0.10,0.13,0.12), RSE_PrevR = c(0.094,0.095,0.05),
        BS_Count = 10000, Boot = FALSE, BMest = 'MDRI.FRR.indep',
        MDRI = c(200,180,180), RSE_MDRI = c(0.05,0.07,0.06),
        FRR = c(0.01,0.009,0.02), RSE_FRR = c(0.2,0.2,0.1), BigT = 730)
```

```
## $Incidence.Statistics
```

```
##    survey Incidence  CI.low   CI.up     RSE RSE.Inf.SS
## 1       1   0.04265  0.0324  0.0529 0.12264    0.05392
## 2       2   0.06774 0.05033 0.08515 0.13111    0.07302
## 3       3   0.04847 0.03961 0.05734 0.09332    0.06625
##
## $Incidence.Difference.Statistics
##   compare      Diff CI.Diff.low CI.Diff.up RSE.Diff RSE.Diff.Inf.SS p.value
## 1  1 vs 2 -0.02509    -0.04529   -0.00489  0.41077         0.21738 0.01492
## 2  1 vs 3 -0.00583    -0.01938    0.00773  1.18669         0.67779 0.39941
## 3  2 vs 1  0.02509     0.00489    0.04529  0.41077         0.21738 0.01492
## 4  2 vs 3  0.01927    -0.00027    0.03880  0.51737         0.30610 0.05326
## 5  3 vs 1  0.00583    -0.00773    0.01938  1.18669         0.67779 0.39941
## 6  3 vs 2 -0.01927    -0.03880    0.00027  0.51737         0.30610 0.05326
##   p.value.Inf.SS
## 1        <0.0001
## 2        0.14011
## 3        <0.0001
## 4        0.00109
## 5        0.14011
## 6        0.00109
##
## $MDRI.CI
##    CI.low   CI.up
## 1 180.400 219.600
## 2 155.304 204.696
## 3 158.832 201.168
##
## $FRR.CI
##   CI.low  CI.up
## 1 0.0061 0.0139
## 2 0.0055 0.0125
## 3 0.0161 0.0239
```

Function `prevcounts`

Function `prevcounts`, while not strictly necessary (and indeed not recommended for final inference on incidence based on real survey data, presumably obtained at great cost and with considerable complex sampling structure) turns counts of:

- Number of subjects tested for HIV
- Number of subjects who tested positive for HIV
- Number of HIV positive subjects tested for recent infections
- Number of HIV positive subjects who tested positive for recent infections

into (point) estimates (and variance) of prevalence of HIV and prevalence of recent infection among HIV the positives. At heart, this is a relatively simple multinomial distribution analysis (trinomial, in the case of complete coverage of recency testing amongst HIV positives) and could be accomplished without any significant innovation directly arising out of the core methods of this package, but function `prevcounts` at least provides a consistent entry point into this analysis, using arguments consistently named to align to the other functions, including appropriate design effects. The most likely use of `prevcounts` is probably indirectly through `inccounts`, but it is provided in user-exposed form for its intuitive supportive value and for recycling into user customisations beyond routine primary incidence estimation. Note that the use of `prevcounts` implies an interpretation of these counts which precludes non-null covariance of the prevalence of HIV and the prevalence of recency.

For a single survey:

```
prevcounts(N = 5000, N_H = 1000, N_testR = 1000, N_R = 70, DE_H = 1.1,
           DE_R = 1.5)
```

```
##   PrevH PrevR  RSE_PrevH RSE_PrevR
## 1   0.2   0.07 0.02966479 0.1411686
```

Note that:

- It is spelled out that in this instance all HIV positive subjects were tested for recent infection

- A design effect is provided for adjusting the variance of the prevalence of HIV

- An independent design effect is provided for adjusting the variance of the prevalence of recent infection amongst HIV positives

- There is no mention of a total sample size in excess of the number of individuals in the underlying survey, on whom there is no HIV status information. A more sophisticated analysis might well use this larger sample size, and additionally account for the frequency, and risk-factor distribution, of missingness, but the conception `prevcounts` does not require this data. The use of design effects, though limited and ultimately problematic, is the appropriate way to insert distributional information beyond the implied two independent binomial distributions for each prevalence.

Input can be provided for two or more surveys in vector form, using the concatenation expression `c()`:

```
prevcounts (N = c(5000,6000), N_H = c(1000,1100), N_testR = c(950,1060),
            N_R = c(100,70), DE_H = c(1.1,1.2), DE_R = c(1.2,1.3))
```

```
##        PrevH      PrevR  RSE_PrevH RSE_PrevR
## 1 0.2000000 0.10526316 0.02966479 0.1036187
## 2 0.1833333 0.06603774 0.02984810 0.1317005
```