Inference with inctools

Avery McIntosh

2016-05-04

The *inctools* package provides two functions for HIV incidence estimation: *incprops* and *inccounts*, and ancillary function *prevcounts*. These functions take as argument user-defined parameters and estimates taken from assays of recent HIV infection and output estimates for inference, incidence difference in the situation of multiple cross-sectional surveys (and possibly multiple assays), and confidence intervals, among other output. See the function help pages for specific examples of running the code. The principal reference for the methods behind this implementation is given by Kassanjee et al. *Epidemiology*, 2012.[1] Further guidance is provided in Kassanjee, McWalter, Welte. *AIDS Research and Human Retroviruses*, 2014.[2]

This vignette will describe the use of necessary functions and some background on methods specific to this package.

prevcounts

Function `prevcounts`, while not strictly necessary for inference on incidence, allows for using function `incprops` while taking into account of design effects, which factor into the relative standard error of prevalence of HIV positivity and recency. `inccounts` has an option for design effects. The function `prevcounts` takes count data from a prevalence survey and produces estimates of prevalences and relative standard errors. Design effects are assumed negligible unless the user specifies otherwise. Note that the input values for design effects are assumed to be known by the user.

For a single survey:

```
library("inctools")
```

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

```
##   PrevH PrevR  RSE_PrevH RSE_PrevR
## 1   0.2  0.07 0.02828427 0.1208896
```

the output for the function is estimated prevalence of HIV given survey counts (5000/1000), estimated proportion of those tested who are defined as recently infected by the assay, and the relative standard error (RSE)–the ratio of standard deviation of the estimator to the estimate, which is itself an estimate of σ/μ.

The number of individuals tested for recency of infection is 1000 (in this example equal to the total number of individuals who tested positive for HIV).

The number of recent infections was calculated by the assay to be 70.

Design effects in the above example are assumed negligible for test of positivity, but are given as 1.1 for the test of recency.

For two or more surveys, the function input is given in vector form using the concatenation expression `c()`.

---

[1]Kassanjee, R., McWalter, T.A., Baernighausen, T. and Welte, A. "A new general biomarker-based incidence estimator." Epidemiology; 2012, 23(5): 721-728; doi:%5B10.1097/EDE.0b013e3182576c07%5D(http://dx.doi.org/10.1097/EDE.0b013e3182576c07); PubMed.

[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; doi:%5B10.1089/aid.2013.0113%5D(http://dx.doi.org/10.1089/aid.2013.0113); PubMed.

```
prevcounts (N = c(5000,5000), N_H = c(1000,1000), N_testR = c(1000,1000),
N_R = c(100,70), DE_H = 1, DE_R = c(1,1.1))
```

```
##   PrevH PrevR  RSE_PrevH  RSE_PrevR
## 1   0.2   0.10 0.02828427 0.09486833
## 2   0.2   0.07 0.02828427 0.12088956
```

Note the construction that survey 1 (the first survey entered) had design effects assumed negligible (=1), while the second survey had an estimated design effect (=1.1). These inputs must be followed sequentially for surveys in multiple entries.

`incprops` and `inccounts`

The functions `incprops` and `inccounts` uses an identical syntax scheme to that shown above for `prevcounts`, and given in the functions' help pages. Both functions take data that has already been calculated about the assay used for testing recency–estimates of false recency rate (FRR–β) and mean duration of recent infection (MDRI–Ω_T) and their respective relative standard errors and the time cut threshold for recency in days (T)–as well a test rejection threshold α (default=0.05) and survey data: proportions (counts, if using function `inccounts`) of HIV positives (PrevH) and positives for recency (PrevR) and their relative standard errors.

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 use input.

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. Run the code given in the examples section at the bottom of the function help page `help("incprops")` for examples of this implementation and output.

The delta method calculation of the estimator variance is described in the second reference, however the delta method approximation of the variance of the difference of incidence estimators has not been described until now. There are three cases that can arise: (1) the same test is used for both assays; (2) the tests are different with independent FRR; (3) the tests are independent with different FRR and different MDRI. The option for assay scheme is specified in both `incprops` and `inccounts` 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.

Case 1: Same test

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

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

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

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

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

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

$$\tag{7}$$

Case 2: FRR independent

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

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

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

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

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

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

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

Case 3: FRR and MDRI independent

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

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

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

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

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

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

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

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

If bootstrapping is use in place of the delta method approximation, the user may also enter in a covariance estimate between testing for recency and testing for positivity to be used in the empirical bootstrap sampler.

The Empirical Bootstrapping Option

Functions `incprops` and `inccounts` both admit the option of bootstrapping. The default multivariate delta method approximation to the variance of the incidence estimator is adequate for inputs that are normal, however the method is still only a linear approximation. Since there are situations where the higher order terms in the Taylor expansion of variance may contribute non-negligibly to the variance estimator, and since it may be the case that both FRR and structured survey data may not be exactly normally distributed, we have programmed an empirical bootstrapping option for users wishing to make non-parametric inference on differences in proportions.

This method is called in `incprops` and `inccounts` by adding the option `Boot = TRUE`. The method is described below.

We start with estimates on parameters that are derived from data. So we have estimates:

- $\widehat{\text{PrevH}}$,
- $\widehat{\text{PrevR}}$,
- $\widehat{\Omega}$,
- $\widehat{\beta}$,

along with their respective estimates of variance.

We resample a multivariate normal distribution with argument `EMPIRICAL = TRUE` to the function `mvrnorm` from the *MASS* package, which coerces the samples such that if we take the mean and variance of any number of samples of Ω_T, say, we will get the exact mean and variance we entered into the multivariate normal sampler as parameters from the distribution. Each unique re-sample for each estimator φ is a B-tuple of values such that the B-tuple mean equals $\widehat{\varphi}$ and its variance exactly equals the variance of the original estimator. Corresponding to each column of this matrix is a single "bootstrap" estimate $\mathbb{I}$, the estimate of incidence.

So we have sample matrix

$$\mathbf{B} = \begin{pmatrix} \text{PrevH}_1 & \text{PrevH}_2 & \text{PrevH}_3 & \cdots & \text{PrevH}_B \\ \text{PrevR}_1 & \text{PrevR}_2 & \text{PrevR}_3 & \cdots & \text{PrevR}_B \\ \Omega_1 & \Omega_2 & \Omega_3 & \cdots & \Omega_B \\ \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_B \end{pmatrix}$$

$$\downarrow \quad\quad \downarrow \quad\quad \downarrow \quad\quad \cdots \quad\quad \downarrow$$

$$\mathbb{I}_1 \quad\quad \mathbb{I}_2 \quad\quad \mathbb{I}_3 \quad\quad \cdots \quad\quad \mathbb{I}_B$$

which is sampled in such a way that, for example, for empirical bootstrap sample Ω_B

$$\sum_{b=1}^{B} \Omega_b = \widehat{\Omega} \quad \text{(the \textit{observed} value of the estimator)}$$

and

$$\frac{1}{B-1} \sum_{b=1}^{B} \Omega_b = s_{\Omega}^2 \quad \text{(the \textit{observed} SD of the estimator)}$$

The bootstrap option produces most of the statistics given by running the functions using a delta method approximation, however the estimate of RSE of the incidence estimator at infinite sample size, which is calculated by letting sample size n tend toward infinity in the limit in the components of the variance of incidence estimator, is not produced.