

# Survey Design (sample size, power, precision)

*Avery McIntosh*

*2016-05-04*

This vignette covers the use of functions `sspower` and `ssprecision`.

## Sample size for power: `sspower`

A common statistical question is whether there is adequate power for a particular study design; that is, does the probability of discovering a statistically significant result rise above a pre-specified threshold, given assumptions on effect size and sample size.

`sspower` takes assumed study parameter values—incidences, prevalences, coverage rates, design effects, and assay characteristics false recency rate (FRR) and mean duration of recent infection (MDRI)—and returns the necessary common sample size of two surveys or the sample size needed for desired power for one survey given that the other survey has already been completed that is required to achieve a given probability of rejecting the statistical test for a null hypothesis that the incidence rates are identical between populations. The function will also return the power of a test of differences under a particular sample size scenario. As supplementary output the function also returns implied statistics such as relative standard error (RSE) from input values on several parameters, their confidence limits, and population counts.

As an example, consider the situation where we wish to discover the power to detect a difference in incidence estimates for a hypothetical scenario where the values of the survey are imagined, respectively, to be incidence 5%, 3% and prevalences at 20%. The assay in this scenario is assumed identical for both surveys. Thus there are unique values for MDRI, the RSE of MDRI, FRR, the RSE of FRR, and time cut T, respectively: 200 days, 5%, 1%, 20%, and 730 days. We envision total coverage, and no survey design effects. Finally, we envision a common study sample size of 5000 persons and we decide on a false positive rate  $\alpha$  of 5%. Note that here function parameter for sample size is listed as `SS = NULL`, or it is omitted in the function call.

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

```
sspower(I1 = 0.05, I2 = 0.03, PrevH1 = 0.20, PrevH2 = 0.20, n1 = 5000, n2 = 5000, alpha = 0.05, Power =
```

```
## $Inc.Difference.Statistics
##   deltaI_Est RSE_deltaI RSE_deltaI.infSS Power Power.infSS CI.low  CI.up
## 1         0.02      0.33          0.052 0.858          >0.99 0.0071 0.0329
##
## $Implied.Incidence.Statistics
##   Survey Given.I RSE_I CI.low CI.up
## 1      1      0.05 0.115  0.039 0.061
## 2      2      0.03 0.147  0.021 0.039
##
## $Implied.MDRI.Statistics
##   Given.MDRI CI.low CI.up
## 1         200 180.4 219.6
##
## $Implied.FRR.Statistics
##   Given.FRR CI.low CI.up
## 1         0.01  0.006 0.014
##
```

```
## $Implied.Subject.Counts
##               Survey.1 Survey.2
## HIV.negative      4000      4000
## HIV.positive      1000      1000
## HIV.post.tested.for.recent 1000      1000
## Recency.test.pos    116        73
```

Output is presented in list form, so individual tables can be called by the user. For example, if the function object is defined as `desired.power`, then the implied subject counts deduced from assumed prevalence and given sample size can be called by entering `desired.power$Implied.Subject.Counts` or `desired.power[[5]]` into the console.

Here the output returns that the power of this particular test given the hypothetical study parameters and sample size is 0.858. In the event that we had an infinite sample size, which of course is impossible, but which is instructive to understand what amount of power is a function of sample size alone and not test characteristics, we would have a power approaching unity.

If instead of power the user wants to know a common sample size for a desired power level, the input parameters are mostly the same as above, except that each variable `n1`, `n1` are to be left unspecified in the function call or both set to "both", and variable `Power` will be set to the desired level. Here we envision the situation where the FRR is independent between tests and will be specified as a tuple in the function. As before there are no design effects or lacking testing coverage. Note also that now the function parameters for sample size is set to `SS = "out"`.

```
sspower(I1 = 0.05, I2 = 0.03, PrevH1 = 0.20, PrevH2 = 0.15, n1 = "both", n2 = "both", alpha = 0.05, Power = 0.858, SS = "out")
```

```
## $Minimum.Common.SS
## [1] 4122
##
## $Inc.Difference.Statistics
##   deltaI_Est RSE_deltaI RSE_deltaI.infSS Power Power.infSS CI.low CI.up
## 1      0.02    0.357      0.063    0.8      >0.99  0.006 0.034
##
## $Implied.Incidence.Statistics
##   Survey Given.I RSE_I CI.low CI.up
## 1      1     0.05 0.124  0.038 0.062
## 2      2     0.03 0.150  0.021 0.039
##
## $Implied.MDRI.Statistics
##   Given.MDRI CI.low CI.up
## 1      200  180.4 219.6
##
## $Implied.FRR.Statistics
##   Given.FRR CI.low CI.up
## 1      0.010 0.0061 0.0139
## 2      0.009 0.0051 0.0129
##
## $Implied.Subject.Counts
##               Survey.1 Survey.2
## HIV.negative      3298      3504
## HIV.positive       824       618
## HIV.post.tested.for.recent 824       618
## Recency.test.pos     95        61
```

The function call outputs that the necessary common study sample size is 4122 persons per study to achieve the desired 80% power given the specified population parameters and assay characteristics.

Finally, imagine the situation where one study has already been completed, and values have been calculated for that study and for the recency assay in use, and the user wishes to find the sample size for the second to-be completed study needed to have a desired power to detect a difference in incidences. The function call for this situation is similar to the previous call, except the survey for which the sample size generation is to be performed will have the sample size parameter in the function call specified as "out". See below for an example of this syntax and output. Again, note that the function parameter for sample size is still set to SS = "out".

```
sspower(I1 = 0.05, I2 = 0.03, PrevH1 = 0.20, PrevH2 = 0.15, n1 = 5000, n2 = "out", alpha = 0.05, Power = 0.80)
```

```
## $Minimum.SS
## [1] 3110
##
## $Inc.Difference.Statistics
##   deltaI_Est RSE_deltaI RSE_deltaI.infSS Power Power.infSS CI.low CI.up
## 1      0.02      0.357      0.052    0.8      >0.99    0.006 0.034
##
## $Implied.Incidence.Statistics
##   Survey Given.I RSE_I CI.low CI.up
## 1      1      0.05 0.115  0.039 0.061
## 2      2      0.03 0.171  0.020 0.040
##
## $Implied.MDRI.Statistics
##   Given.MDRI CI.low CI.up
## 1      200  180.4 219.6
##
## $Implied.FRR.Statistics
##   Given.FRR CI.low CI.up
## 1      0.01 0.0059 0.0141
##
## $Implied.Subject.Counts
##                               Survey.1 Survey.2
## HIV.negative                 4000      2644
## HIV.positive                 1000      466
## HIV.post.tested.for.recent    1000      466
## Recency.test.pos              116       47
```

The function now returns the necessary sample size for survey 2 that will generate power of 80%, given the hypothetical population parameters for survey 2 and the observed parameter estimates for survey 1 and the assay characteristics MDRI, FRR, and their respective relative standard errors and time cut point T.

## Sample size for precision: `ssprecision`

This function summarizes performance of a recent infection test into a standard error of the incidence estimate given estimated test properties and the prevalence and incidence in a hypothetical context. Alternatively the function can return the sample size necessary for a given level of estimator precision. Up to two parameters can be given as tuple vectors, with the input parameter `step` specifying the number of points analyzed between the endpoints of the given tuple. This specification will yield output for each value in the step for the output parameters that take as argument one of the varying inputs. See the second and third example below for an illustration of this output.

The function call below returns the necessary sample size to have RSE of the incidence estimator equal to 25%, given a hypothetical prevalence, coverage rate, and assay tests parameter estimates. Note here the design effects for the test of HIV positivity are set to 1.1 and that sample size is set to `n = "out"`.

```
ssprecision(I = 0.015, RSE_I = 0.25, PrevH = 0.2, CR = 1, MDRI = 200, RSE_MDRI = 0.05, FRR = 0.01, RSE_I
```

```
## $sample.size
## [1] 3634
##
## $Prev.HIV.and.recent
## [1] 0.00833
##
## $Prev.HIV.and.nonrecent
## [1] 0.19167
##
## $RSE.I.inf.sample
## [1] 0.07606
##
## $RSE.PrevH
## [1] 0.0348
##
## $RSE.PrevR
## [1] 0.03558
```

The function call returns that the necessary sample size for the desired precision in the incidence estimator is 3634 persons. Prevalence of being HIV positive and recently infected is 0.8%, prevalence of being HIV positive and nonrecently infected is 19%. The RSE of the incidence estimator at infinite sample size is 7.6%, and the RSE of prevalence of positivity for HIV is 3.4% and the RSE of prevalence of recency is 3.55%.

Next consider the function call for a tuple of prevalence of HIV positivity values and incidence estimates that range in value between 10 and 20% and 1.5 and 2% respectively. The function now outputs a range of values 3 values for each output parameter that is dependent upon these two varying quantities, as well as the normal output from a single parameter function call when the output does not depend on the varying parameters. Note again that sample size is set to `n = "out"`.

```
ssprecision(I = c(0.015,0.02), RSE_I = 0.25, PrevH = c(0.10,0.20), CR = 1, MDRI = 200, RSE_MDRI = 0.05,
```

```
## $sample.size
##          PrevH = 0.1 PrevH = 0.15 PrevH = 0.2
## I = 0.015      2660      2660      2660
## I = 0.0175     2547      2547      2547
## I = 0.02       2489      2489      2489
##
## $Prev.HIV.and.recent
##          PrevH = 0.1 PrevH = 0.15 PrevH = 0.2
## I = 0.015     0.00813     0.00813     0.00813
## I = 0.0175     0.00936     0.00936     0.00936
## I = 0.02       0.01045     0.01045     0.01045
##
## $Prev.HIV.and.nonrecent
##          PrevH = 0.1 PrevH = 0.15 PrevH = 0.2
## I = 0.015     0.09187     0.09187     0.09187
## I = 0.0175     0.14064     0.14064     0.14064
```

```
## I = 0.02      0.18955      0.18955      0.18955
##
## $RSE.I.inf.sample
##      PrevH = 0.1 PrevH = 0.15 PrevH = 0.2
## I = 0.015      0.05583      0.05583      0.05583
## I = 0.0175     0.06033      0.06033      0.06033
## I = 0.02       0.06549      0.06549      0.06549
##
## $RSE.PrevH
##      PrevH = 0.1 PrevH = 0.15 PrevH = 0.2
## I = 0.015      0.05817      0.05817      0.05817
## I = 0.0175     0.04717      0.04717      0.04717
## I = 0.02       0.04009      0.04009      0.04009
##
## $RSE.PrevR
##      PrevH = 0.1 PrevH = 0.15 PrevH = 0.2
## I = 0.015      0.02061      0.02061      0.02061
## I = 0.0175     0.02975      0.02975      0.02975
## I = 0.02       0.03817      0.03817      0.03817
```

So for example in the first listed output, the sample size necessary for a given precision in the incidence estimator, the output shows that for hypothetical values of PrevH = 0.15 and I = 0.0175, midway in the step between 10 and 20% and 1.5 and 2% respectively, the necessary sample size for RSE\_I = 0.25 is 2547 persons enrolled.

Finally, for calculating the RSE of incidence over a range of 5 values of prevalence of positivity with a sample size of 5000, the user will set the sample size parameter to a given value, here n = 5000. The precision is set to output with option RSE\_I = "out".

```
ssprecision(I = 0.017, RSE_I = 'out', PrevH = c(0.10,0.20), CR = 1, MDRI = 211, RSE_MDRI = 0.05, FRR = 0)
```

```
## $RSE_I
##      PrevH  RSE_I
## 1 PrevH = 0.1 0.16868
## 2 PrevH = 0.125 0.17352
## 3 PrevH = 0.15 0.17885
## 4 PrevH = 0.175 0.18470
## 5 PrevH = 0.2 0.19112
##
## $Prev.HIV.and.recent
##      PrevH Prev.HIV.and.recent
## 1 PrevH = 0.1      0.00947
## 2 PrevH = 0.125      0.00945
## 3 PrevH = 0.15      0.00944
## 4 PrevH = 0.175      0.00943
## 5 PrevH = 0.2      0.00942
##
## $Prev.HIV.and.nonrecent
##      PrevH Prev.HIV.and.nonrecent
## 1 PrevH = 0.1      0.09053
## 2 PrevH = 0.125      0.11555
## 3 PrevH = 0.15      0.14056
## 4 PrevH = 0.175      0.16557
## 5 PrevH = 0.2      0.19058
```

```

##
## $RSE.I.inf.sample
##      PrevH RSE.I.inf.sample
## 1  PrevH = 0.1      0.05363
## 2 PrevH = 0.125      0.05557
## 3  PrevH = 0.15      0.05824
## 4 PrevH = 0.175      0.06166
## 5  PrevH = 0.2      0.06585
##
## $RSE.PrevH
##      PrevH RSE.PrevH
## 1  PrevH = 0.1  0.04243
## 2 PrevH = 0.125 0.03742
## 3  PrevH = 0.15 0.03367
## 4 PrevH = 0.175 0.03071
## 5  PrevH = 0.2  0.02828
##
## $RSE.PrevR
##      PrevH RSE.PrevR
## 1  PrevH = 0.1  0.01383
## 2 PrevH = 0.125 0.01748
## 3  PrevH = 0.15 0.02113
## 4 PrevH = 0.175 0.02479
## 5  PrevH = 0.2  0.02845

```