

PF Package

Examples for Stratified Designs

Center for Veterinary Biologics - Statistics Section

July 2019

Contents

1	Introduction	1
1.1	Model assumptions and criticisms	2
1.2	Alternatives	3
2	Theory	4
2.1	Notation	4
2.2	Mantel-Haenszel (MH) estimator	4
2.3	Gart-Nam (GN) estimator	5
3	Examples	6
3.1	An example with 8 strata	6
3.2	Examples with heterogeneity	9
3.3	An example with 9 strata	12
3.4	Examples with only two strata	14
3.5	An example with a non-informative stratum	16
3.6	A large sample example	18
4	Future work	20
	References	20

1 Introduction

Consider a prospective, parallel-group, placebo-controlled, randomized, blinded clinical trial with a single binary endpoint (e.g., positive or negative for disease). In veterinary biologics research, an efficacy trial for a vaccine may often follow a vaccination-challenge protocol. For simplicity, suppose that there are only two treatment groups, vaccinate and (placebo) control. Suppose further that there is also a stratification variable with K levels, crossed with treatment group. This stratification variable could be cohort, litter, housing unit, site, etc. Animals from the two groups are commingled within strata. Define the *risk ratio* (RR) as the fraction of vaccinates positive for disease divided by the fraction of control animals positive for disease. The *prevented fraction* (PF) is $1 - RR$.

This vignette will consider the estimation of the risk ratio (or prevented fraction) particularly in the case where the number of animals per stratum is small, the number of strata is small, and the results are *sparse*. By sparse we mean that in the cross tabulation of results, many low or zero counts show up in the cells of the $K \times 2 \times 2$ tables. This occurs, for instance, when nearly all vaccinates are negative for disease and nearly all controls are positive. In such cases, the methods of (Gart and Nam 1988; Mantel and Haenszel 1959) may be particularly useful. Functions using these two methods are included in the PF package as `RRmh()` and `RRstr()`, respectively.

Good general references on the class of problems considered here include (Agresti 2002) and (Lachin 2011), although neither of these covers the Gart-Nam method. We are grateful to colleagues in the CVB Statistics Section for helpful discussion.

1.1 Model assumptions and criticisms

Both the Mantel-Haenszel and Gart-Nam methods rely on the asymptotic normality of the $\log(RR)$, % which I denote as θ , although the Gart-Nam procedure includes a skewness correction to mitigate the use of this assumption. Nonetheless, the use of an asymptotic result could be questioned for the case of small samples. Unfortunately, all of the existing work on exact methods for this problem is focused on the odds ratio, not the risk ratio. These include StatXact's `Proc Stratify` as well as exact (conditional) logistic regression methods provided by either SAS's `Proc Logistic` or Cytel's `LogXact`.

Work on exact analysis of risk difference and relative risk models [for] $K \times 2 \times 2$ tables appears nonexistent" (Hirji 2006,p. ~274). Exact methods aside, many other alternative analysis procedures, such as logistic regression, also rely on asymptotics (McCulloch et al. 2008).

Secondly, both the Mantel-Haenszel and Gart-Nam methods make the assumption of *homogeneity*, that there is no treatment group-by-stratum interaction. Specifically, they assume that the prevalence of disease in the placebo group varies across strata, but that the risk ratio is the *same* for all strata. For small, sparse data sets, it can be difficult to assess the plausibility of this assumption. Tests for homogeneity often lack power for such data sets, except for the most egregious cases, which can usually be detected by visual inspection.

When the assumption is false, a question is raised of what is meant by an overall risk ratio, and why should you be estimating it? It is plausible that the assumption is false when a large variation in disease prevalence in the placebo group across strata is evident. The latter may suggest variation in challenge exposure, perhaps due to varying environmental conditions. In that situation the apparent efficacy of the vaccine may be expected to vary.

A third criticism of the Gart-Nam and Mantel-Haenszel approaches is that neither is able to accomodate other covariates, unlike the logistic regression approach. If there are two crossed stratification variables (e.g., cohort and pen), one could replace them with a composite stratification variable consisting of all combinations of levels of the two original ones. Nested stratification variables could be partially accomodated by using only the lowest level of nesting (and ensuring that all levels are uniquely labeled).

1.2 Alternatives

A few words can be said about alternative approaches to estimation. Many of these rely on generalized linear models (GLMs) that fall into the category of logistic regression models. When accounting for strata in some way, these models typically imply that there is a common odds ratio for all strata, but that the prevalence of disease among control animals may vary between strata.

A fixed effect GLM would include a separate parameter for the baseline odds of disease in each stratum. If the link is logit, the model can estimate a common odds ratio but not a common risk ratio (as could be done by transformation, i.e., by using the delta method, when there are no stratum-specific covariates).

To estimate the risk ratio, the log link would need to be used, as suggested by some authors in other contexts.¹

Another challenge for logistic regression is the case of complete separation, when it occurs. (An example of complete separation is when all controls are positive for disease and all vaccinates are negative for disease, in all strata. In this case, the covariate is able to perfectly separate positives from negatives.) Complete separation results in a non-existent maximum likelihood for the logistic regression model, although there are mitigation strategies (Heinze and Schemper 2002).

Binomial GLMs may be extended relaxing the assumption that the dispersion parameter is unity. Quasi-likelihood (Wedderburn 1974) or extended quasi-likelihood (McCullagh and Nelder 1989, Section 10.4) models provide a means to estimate the dispersion parameter.

A fully parametric approach to the same end would be a hierarchical mixture model such as the beta-binomial. Such models could be used to estimate a risk ratio by transformation, with the understanding that the estimate is a kind of average over the strata. However, when fitting such models, one withholds certain information about the study design, namely the assignment of subjects to strata.

Random effects models are sometimes used in a similar fashion, but they are different.² In quasi-likelihood models, the dispersion is a weight of the mean. In a random effect model, the random effect is a modification to the logit transformed mean. It would be questionable to estimate a risk ratio from such a model.

¹E.g., See (Wacholder 1986) for a model using the log link in a binomial GLM or quasi-likelihood GLM with binomial variance function. See (Carter et al. 2005; Zou 2004) for a model using the log link with a robust Poisson variance function. Such efforts are not always successful, because unlike the logit link, which constrains estimation of the binomial parameter to the appropriate parameter space, the log link does not (Yelland et al. 2011).

²The term *random effects model* these days has come to mean tacking on an additive normally distributed random variable to whatever likelihood is supported by the sampling distribution. This practice is also done for GLMs, where they are called generalized linear mixed models in emulation of linear mixed models, and the prevalence of software has encouraged their widespread use without much consideration for their plausibility or suitability. Typically stratum would be included as a random but not fixed effect in the linear predictor. There are also technical issues associated with such models (Agresti and Hartzel 2000).

Table 1: Notation for j th stratum

Group	Positives Animals	Negatives Animals	Total Animals
Vaccinate	y_{1j}	$n_{1j} - y_{1j}$	n_{1j}
Placebo	y_{2j}	$n_{2j} - y_{2j}$	n_{2j}

2 Theory

2.1 Notation

In stratum j , for $1 \leq j \leq K$, define the following notation. Suppose that π_{1j} is the probability for a vaccinate to be positive and π_{2j} is the probability for a control animal to be positive. Denote the number of animals positive and negative for disease as in Table 1.

Then let $p_{1j} = y_{1j}/n_{1j}$ and $p_{2j} = y_{2j}/n_{2j}$ be estimates for π_{1j} and π_{2j} , respectively. An estimate of the risk ratio in the j th stratum is then $\hat{RR}_j = p_{1j}/p_{2j}$.

2.2 Mantel-Haenszel (MH) estimator

The (Mantel and Haenszel 1959) estimator for RR , using the asymptotic variance formula for sparse asymptotics (Greenland and Robins 1985), is as follows.

$$\hat{RR} = \frac{\sum_{j=1}^K y_{1j}n_{2j}/(n_{1j} + n_{2j})}{\sum_{j=1}^K y_{2j}n_{1j}/(n_{1j} + n_{2j})}. \quad (1)$$

$$\hat{Var}(\log \hat{RR}) = \frac{\sum_{j=1}^K (n_{1j}n_{2j}(y_{1j} + y_{2j}) - y_{1j}y_{2j}(n_{1j} + n_{2j})) / (n_{1j} + n_{2j})^2}{\left(\sum_{j=1}^K \frac{y_{1j}n_{2j}}{n_{1j} + n_{2j}}\right) \left(\sum_{j=1}^K \frac{y_{2j}n_{1j}}{n_{1j} + n_{2j}}\right)}. \quad (2)$$

Note that if $y_{1j} = 0$ or $y_{2j} = 0$ for all strata, then there will be a division by zero error. In other words, if all the vaccinates are negative for disease and/or all the controls are negative for disease, we will not be able to use the Mantel-Haenszel estimator. Do **not** attempt to remedy this problem by replacing the zero cells with tiny constants (Agresti and Hartzel 2000).

Greenland and Robins (1985) find that the Mantel-Haenszel estimator is consistent for both the case of sparse strata where the number of strata is assumed increasing, and the case of a limited number of strata where the stratum size is assumed increasing. In the latter case, however, the MH estimator is less efficient than maximum likelihood. They recommend that the MH estimator only be used for sparse data.

In that situation, Agresti and Hartzel (2000) find that the Mantel-Haenszel estimator is preferred to maximum likelihood estimators from generalized linear models, due to the latter's known bias for sparse data.

A broader perspective on the Cochran-Mantel-Haenszel framework, which is usually discussed in terms of odds ratio estimation rather than risk ratio estimation, can be found in Kuritz et al. (1988).

The above MH estimator is implemented in SAS `Proc Freq`, according to the documentation for SAS 9.3. It is also implemented as the `mhgr()` function in Prof.~Frank Harrell, Jr.’s `Hmisc` R package; its interface differs from the one provided for `RRmh()` in the `PF` package discussed below.

SAS’s `Proc Freq` also provides a weighted least squares estimate that it calls `logit` (Lachin 2011; Stokes et al. 2000). The latter estimator is **not** recommended due to “large and erratic bias” among other reasons (Greenland and Robins 1985).

2.3 Gart-Nam (GN) estimator

The Gart and Nam (1988) estimator is a skew-corrected version of the Gart (1985) estimator, which is related to the Radhakrishna (1965) test against the null hypothesis $RR = 1$. The skew correction mitigates the use of an asymptotic variance for the $\log RR$, especially for unblanced designs or when the π ’s are near zero or one.

There is no closed form expression for the Gart (1985) estimator. An iterative algorithm to calculate it can be described as follows:

- Assume that ϕ is the common risk ratio across all strata.
- Write the likelihood as the product of two binomial distributions, with parameters π_{1j} and π_{2j} for each stratum j .
- Set $\pi_{1j} = \phi\pi_{2j}$. Now there are $K + 1$ unknowns, ϕ and the π_{2j} ’s.
- Set the nuisance scores (partial derivatives of the log likelihood with respect to the K parameters π_{2j}) equal to zero. This results in a set of K quadratic equations to solve, for a given ϕ , for the maximum likelihood estimators of the π_{2j} ’s.
- The distribution of the score for ϕ is asymptotically normal with mean zero, and variance $1/((1 - \hat{\pi}_{1j})/(n_{1j}\hat{\pi}_{1j}) + (1 - \hat{\pi}_{2j})/(n_{2j}\hat{\pi}_{2j}))$. Combine this variance expression with the nuisance equations and solve iteratively.
- Finally, apply the Cornish-Fisher skewness correction to the confidence interval limits.

Sadly, the Gart-Nam method for stratified data is not usually incorporated in popular software packages.

(Several software packages incorporate the analogous Gart-Nam estimate for a *single* stratum, such as StatsDirect, NCSS [Number Crunching Statistical Systems], and the R package `pairwiseCI`.) Happily the `PF` package includes an implementation of the Gart-Nam method for stratified designs in the `RRstr()` function.

3 Examples

As a source of examples, we will use either literature (with citations included) or mock examples based on our experience.

When comparing results with those calculated by pooling the strata, the pooled estimator we use is the ‘exact’ interval based on inverting the two one-sided score tests (TOSST). The function `RRtosst()` in the `PF` package implements this procedure. See the `PF` vignette for more details. The pooled estimator can be hazardous and result in, e.g., Simpson’s paradox type phenomena, where there is an apparent effect in the opposite direction of the true effect.

Even when there is no heterogeneity, the use of a pooled estimator can be misleading if the ratio of sample sizes between the two treatment groups is different among strata, e.g., (Gart 1962).

We use it here to illustrate the consequences of ignoring strata on the estimates.

First, let’s load the `PF` package.

```
library(PF)
```

In the examples presented below, I input the data matrix directly as a summary table. Both the `RRmh()` and `RRstr()` functions also allow data input by *formula*. See the `PF` vignette and package help files for more details. In that case, the functions also produce the data matrix.

3.1 An example with 8 strata

In this first example, I will flesh out the details of the usage of the R functions in greater detail than in the remaining examples. The data within the strata are sparse and have small sample sizes:

```
input.mtx <- matrix(c(0, 3, 1, 3, 1, 3, 3, 3, 0, 2, 2, 2, 1, 3, 2, 3, 2, 3,
  3, 3, 1, 2, 3, 3, 0, 2, 3, 3, 0, 3, 2, 2), 8, 4, byrow = TRUE)
```

The rows of the matrix represent the eight strata.

The four columns are, from left to right, the number of vaccinates positive for disease, the total number of vaccinates, the number of controls positive for disease, and the total number of controls. A tabulation of the data (where each row is a stratum) is found in Table 2.

First, let’s attempt to (informally) assess homogeneity by calculating the *PFs* within each strata.

```
rr.by.stratum <- (input.mtx[, 1] / input.mtx[, 2]) /
  (input.mtx[, 3] / input.mtx[, 4])
round(1 - rr.by.stratum, 2)
```

```
## [1] 1.00 0.67 1.00 0.50 0.33 0.50 1.00 1.00
```

	Positive vaccinates	Total vaccinates	Positive controls	Total controls
1	0	3	1	3
2	1	3	3	3
3	0	2	2	2
4	1	3	2	3
5	2	3	3	3
6	1	2	3	3
7	0	2	3	3
8	0	3	2	2

Table 2: An example with 8 strata

Four of the eight strata have PF s of one; the other four have PF s of $2/3$, $1/2$, $1/2$, and $1/3$. It is hard to know whether the apparent variability is an artifact of the small sample sizes (2 or 3 animals per group per stratum), which restricts the attainable values of PF . Inidentally, the above calculation of the stratum-specific PF s is automatically done by the `RRstr()` function, as I will show at the end of this section.

How much variation exists in the prevalence of disease among control animals?

```
control.prev <- input.mtx[, 3] / input.mtx[, 4]
round(control.prev, 2)
```

```
## [1] 0.33 1.00 1.00 0.67 1.00 1.00 1.00 1.00
```

Six of the eight strata (75%) have all controls affected by disease; there is only one stratum each with prevalences of 0.33 and 0.67.

As is typical with this kind of data, the question of homogeneity cannot be resolved with any satisfaction. Although homogeneity could be questioned, we will proceed by assuming it. Let's calculate the Mantel-Haenszel estimate of the prevented fraction:

```
RRmh(Y = input.mtx)
```

```
##
## PF
## 95% interval estimates
##
##    PF    LL    UL
## 0.714 0.394 0.865
```

The output includes the 95% confidence intervals, with LL as the lower limit and UL as the upper limit. Here is the Gart-Nam estimate, both with and without the skew-correction:

```
a <- RRstr(Y = input.mtx)
a
```

```
##
## Test of homogeneity across clusters
```

```
##
## stat  6.15
## df    7
## p     0.522
##
## PF
## 95% interval estimates
##
##          PF    LL    UL
## starting 0.605 0.311 0.773
## mle      0.721 0.483 0.876
## skew corr 0.721 0.479 0.889
```

The function first returns the results of the score-based χ^2 -test for homogeneity (Gart and Nam 1988). Next, the starting (initial) value for the prevented fraction estimate is given, followed by the maximum likelihood estimate (MLE) and skew corrected (skew corr) estimate. As before, the 95% confidence interval limits are also included. In this case, the skew correction does not affect the confidence interval much. The GN interval is less wide (less conservative) than the Mantel-Haenszel estimate's. A case where the GN interval is more conservative than the MH interval will be shown in Subsection 3.3.

Finally, note that all the information displayed by the print method of `RRstr()`, and more, can be extracted from the object `a`. Particularly useful are the stratum-specific estimates of the risk ratio, which is accessed through the data matrix:

```
a$y
```

```
##          R.obs
## [1,] 0 3 1 3 0.0000000
## [2,] 1 3 3 3 0.3333333
## [3,] 0 2 2 2 0.0000000
## [4,] 1 3 2 3 0.5000000
## [5,] 2 3 3 3 0.6666667
## [6,] 1 2 3 3 0.5000000
## [7,] 0 2 3 3 0.0000000
## [8,] 0 3 2 2 0.0000000
```

If you prefer to look at the *PF*s, you may use the following calculation:

```
round(1 - a$y[, 5], 2)
```

```
## [1] 1.00 0.67 1.00 0.50 0.33 0.50 1.00 1.00
```

which is equivalent to the calculation of the stratum-specific *PF*s shown earlier.

3.2 Examples with heterogeneity

The next example provides a clearer illustration of inhomogeneity. Here are the data:

```
input.mtx <- matrix(c(0, 3, 3, 3, 0, 1, 1, 1, 1, 2, 1, 2, 2, 3, 2, 2, 1, 1,
  1, 1, 1, 3, 1, 2, 0, 3, 2, 3, 0, 4, 4, 5), 8, 4, byrow = TRUE)
```

A tabulation of the data (where each row is a stratum) is found in Table 3.

	Positive vaccinates	Total vaccinates	Positive controls	Total controls
1	0	3	3	3
2	0	1	1	1
3	1	2	1	2
4	2	3	2	2
5	1	1	1	1
6	1	3	1	2
7	0	3	2	3
8	0	4	4	5

Table 3: An example with inhomogeneity

The prevented fractions for each stratum are as follows.

```
rr.by.stratum <- (input.mtx[, 1] / input.mtx[, 2]) /
  (input.mtx[, 3] / input.mtx[, 4])
round(1 - rr.by.stratum, 2)
```

```
## [1] 1.00 1.00 0.00 0.33 0.00 0.33 1.00 1.00
```

In half of the litters, complete protection of vaccinates is evident ($PF = 1$), in a quarter there is partial protection ($PF = 1/3$), and in the other quarter there is no protection at all ($PF = 0$). The pooled estimate of prevented fraction is

```
RRtosst(apply(input.mtx, 2, sum))
```

```
##
## PF
## 95% interval estimates
##
##    PF    LL    UL
## 0.683 0.315 0.887
```

Here are the MH and GN estimates.

```
RRmh(Y = input.mtx)
```

```
##
## PF
## 95% interval estimates
```

```
##
##      PF      LL      UL
## 0.710 0.317 0.877
RRstr(Y = input.mtx)

##
## Test of homogeneity across clusters
##
## stat  9.49
## df     7
## p      0.219
##
## PF
## 95% interval estimates
##
##              PF      LL      UL
## starting  0.623 0.2545 0.810
## mle       0.679 0.2957 0.860
## skew corr 0.679 0.0384 0.875
```

The MH point estimate is slightly higher than the GN and pooled estimates. On the other hand, the confidence interval width of the MH and pooled estimates are close. Note how different the confidence intervals are for the GN estimate, with and without the skewness correction, in contrast to the last example.

In this example, the assumption of homogeneity is likely to be unconvincing. Thus, any of the estimates reported above could be questioned. Note that the χ^2 -test for homogeneity does not reject the null hypothesis at the 0.05 level, illustrating its lack of power for small samples.

Now let's turn to another example where the inhomogeneity is so severe that even the χ^2 -test will detect it. The strata represent age groups; thus it is questionable to even treat this factor as a stratification variable.

```
input.mtx2 <- matrix(c(9, 10, 4, 5, 1, 10, 5, 5), 2, 4, byrow = TRUE)
```

A tabulation of the data (where each row is a stratum) is found in Table 4.

	Positive vaccinates	Total vaccinates	Positive controls	Total controls
1	9	10	4	5
2	1	10	5	5

Table 4: An example with severe inhomogeneity

The prevented fractions for each stratum are as follows.

```
rr.by.stratum <- (input.mtx2[, 1] / input.mtx2[, 2]) /
  (input.mtx2[, 3] / input.mtx2[, 4])
round(1 - rr.by.stratum, 2)
```

```
## [1] -0.12  0.90
```

In one stratum the treatment is arguably harmful and in the other it is quite protective. Most statisticians would proceed to analyze the strata separately. But, for the sake of illustrating the test for homogeneity, here is the GN estimate.

```
RRstr(Y = input.mtx2)
```

```
##
## Test of homogeneity across clusters
##
## stat  11
## df     1
## p      0.000891
##
## Heterogeneity may be present
##
## PF
## 95% interval estimates
##
##          PF      LL      UL
## starting 0.0978 0.0532 0.140
## mle      0.1111 -0.3206 0.693
## skew corr 0.1111 -0.2868 0.701
```

Look at the p-value for the homogeneity test. The inferences for the *PF* are not valid. There is no single *PF*; the efficacy of the vaccine depends on the age at which vaccination and challenge occur.

A reckless person might compute the pooled estimate,

```
RRtosst(apply(input.mtx2, 2, sum))
```

```
##
## PF
## 95% interval estimates
##
##      PF      LL      UL
## 0.4444 0.0372 0.7000
```

or the MH estimate, which is slightly less conservative:

```
RRmh(Y = input.mtx2)
```

```
##
```

```
## PF
## 95% interval estimates
##
##      PF      LL      UL
## 0.4444 0.0516 0.6746
```

Statistical significance is barely achieved; there would be no sign of trouble unless the individual stratum results were also reported.

3.3 An example with 9 strata

In this example, the strata are litters.

```
input.mtx <- matrix(c(2, 3, 3, 3, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 2, 3,
  3, 3, 4, 4, 4, 4, 3, 3, 3, 3, 3, 4, 4, 4, 1, 2, 3, 3), 9, 4,
  byrow = TRUE)
```

A tabulation of the data (where each row is a stratum) is found in Table 5.

	Positive vaccinates	Total vaccinates	Positive controls	Total controls
1	2	3	3	3
2	2	3	3	3
3	3	3	3	3
4	3	3	3	3
5	2	3	3	3
6	4	4	4	4
7	3	3	3	3
8	3	4	4	4
9	1	2	3	3

Table 5: An example with 9 strata

The prevented fractions for each stratum are as follows.

```
rr.by.stratum <- (input.mtx[, 1] / input.mtx[, 2]) /
  (input.mtx[, 3] / input.mtx[, 4])
round(1 - rr.by.stratum, 2)
```

```
## [1] 0.33 0.33 0.00 0.00 0.33 0.00 0.00 0.25 0.50
```

```
summary(1 - rr.by.stratum)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.0000  0.2500  0.1944  0.3333  0.5000
```

The prevented fractions are low, ranging from zero to 0.50. The prevalence of disease in control animals is 1.00 in all strata. The pooled estimate of PF is

```
RRtosst(apply(input.mtx, 2, sum))
```

```
##
## PF
## 95% interval estimates
##
##      PF      LL      UL
## 0.1786 0.0416 0.3689
```

and the MH estimate of prevented fraction is

```
RRmh(Y = input.mtx)
```

```
##
## PF
## 95% interval estimates
##
##      PF      LL      UL
## 0.1831 0.0239 0.3164
```

which is close to the pooled estimate. However, the GN estimate is

```
RRstr(Y = input.mtx)
```

```
##
## Test of homogeneity across clusters
##
## stat  5.71
## df     8
## p     0.679
##
## PF
## 95% interval estimates
##
##              PF      LL      UL
## starting  0.174   0.117 0.227
## mle       0.174 -11.098 0.356
## skew corr 0.174 -15.307 0.350
```

Although all the point estimates are close to each other, the GN estimate fails to achieve *statistical significance* while the MH and pooled estimates barely do. The point is that in this example GN is much more conservative than MH and the pooled estimate, in contrast to the example in Subsection 3.1.

3.4 Examples with only two strata

In the first such example, males and females were segregated in different rooms:

```
input.mtx <- matrix(c(1, 3, 4, 4, 0, 6, 6, 6), 2, 4, byrow = TRUE)
```

A tabulation of the data (where each row is a stratum) is in Table 6.

	Positive vaccinates	Total vaccinates	Positive controls	Total controls
1	1	3	4	4
2	0	6	6	6

Table 6: First example with 2 strata

The individual prevented fractions by rooms are 0.67 and 1.00, respectively. Here are the MH and GN estimators of prevented fraction.

```
RRmh(Y = input.mtx)
```

```
##
## PF
## 95% interval estimates
##
##      PF      LL      UL
## 0.879 0.301 0.979
```

```
RRstr(Y = input.mtx)
```

```
##
## Test of homogeneity across clusters
##
## stat  2.25
## df    1
## p     0.134
##
## PF
## 95% interval estimates
##
##              PF      LL      UL
## starting  0.833 -1.932 0.991
## mle       0.889  0.565 0.980
## skew corr 0.889  0.574 0.993
```

The point estimates are nearly identical, but the MH confidence interval is much wider (more conservative) than the GN confidence interval. For comparison, the pooled estimate is

```
RRtosst(apply(input.mtx, 2, sum))
```

```
##
```

```
## PF
## 95% interval estimates
##
##      PF      LL      UL
## 0.889 0.518 0.997
```

This is a bit more conservative than the GN estimate, but not as conservative as the MH estimate.

Here is another 2-strata example; the strata represent cohorts.

```
input.mtx2 <- matrix(c(1, 8, 1, 7, 1, 8, 4, 7), 2, 4, byrow = TRUE)
```

A tabulation of the data (where each row is a stratum) is in Table 7.

	Positive vaccinates	Total vaccinates	Positive controls	Total controls
1	1	8	1	7
2	1	8	4	7

Table 7: Second example with 2 strata

In the first cohort, the challenge was very weak, as indicated by the low prevalence of disease in the placebo animals. However, the vaccinates also had one animal affected by disease! Here are estimates of prevented fraction for each stratum:

```
rr.by.stratum <- (input.mtx2[, 1] / input.mtx2[, 2]) /
  (input.mtx2[, 3] / input.mtx2[, 4])
1 - round(rr.by.stratum, 2)
```

```
## [1] 0.12 0.78
```

Homogeneity could certainly be questioned here. If we blithely proceed, here are the MH and GN estimates of prevented fraction.

```
RRmh(Y = input.mtx2)
```

```
##
## PF
## 95% interval estimates
##
##      PF      LL      UL
## 0.650 -0.529 0.920
```

```
RRstr(Y = input.mtx2)
```

```
##
## Test of homogeneity across clusters
##
## stat 0.777
## df   1
```

```
## p      0.378
##
## PF
## 95% interval estimates
##
##          PF      LL      UL
## starting 0.675 -0.860 0.943
## mle      0.671 -0.208 0.915
## skew corr 0.671 -0.245 0.943
```

Neither estimate achieves *statistical significance*. The pooled estimate is

```
RRTosst(apply(input.mtx2, 2, sum))
```

```
##
## PF
## 95% interval estimates
##
##      PF      LL      UL
## 0.650 -0.570 0.946
```

This is comparable to the MH estimate. The GN estimate has a narrower (less conservative) confidence interval.

3.5 An example with a non-informative stratum

In this example, we have three strata representing housing units. Piglets were housed by farrowing date, so the strata also represent cohorts staggered in time. The outcome was mortality. Here is the data matrix:

```
input.mtx <- matrix(c(0, 5, 0, 4, 0, 5, 1, 3, 0, 2, 4, 6), 3, 4,
  byrow = TRUE)
```

A tabulation of the data (where each row is a stratum) is found in Table 8.

	Positive vaccinates	Total vaccinates	Positive controls	Total controls
1	0	5	0	4
2	0	5	1	3
3	0	2	4	6

Table 8: An example with a non-informative stratum

The allocation of animals to strata is unbalanced due both to dropouts and human error. In the first stratum, no vaccinates and no controls were affected by disease; the risk ratio for that stratum is inestimable. This is an example of a so-called *non-informative* table, since we cannot learn anything about the stratum-specific protection offered by the vaccine from

this stratum. The risk ratios in the other two strata are zero. The prevalences of disease in the control animals in each stratum are zero, 1/3, and 2/3, respectively.

If we attempt to estimate the prevented fraction using the Mantel-Haenszel estimator, we obtain

```
RRmh(Y = input.mtx)
```

```
##
## PF
## 95% interval estimates
##
##   PF   LL   UL
##   1 NaN   1
```

The lower confidence limit is inestimable due to the fact that no vaccinates were positive for disease in any stratum. The Mantel-Haenszel estimator cannot be used with this data, regardless of whether we choose to retain or ignore the *non-informative* stratum.

The Gart-Nam estimator breaks down if we include the *non-informative stratum*: if you were to type `RRstr(Y = input.mtx)` you would receive an error message “Error in `zd[is.na(zd)] <- 2 * zd[!is.na(zd)]` : replacement has length zero.” At present the function does not have a graceful exit when an error occurs. Now if we *remove* the *non-informative* stratum, here is what we get.

```
input.mtx2 <- input.mtx[-1,]
RRstr(Y = input.mtx2)
```

```
##
## Homogeneity test not possible because MLE = 1
##
## PF
## 95% interval estimates
##
##           PF      LL UL
## starting   1 0.0000  1
## mle        1 0.0617  1
## skew corr  1 0.0636  1
```

The homogeneity test cannot be carried out. Although the prevented fraction is unity, the confidence interval is very wide.

Now suppose that in addition to mortality we include pigs showing clinical signs of disease as positive for *morbidity*. If so, it turns out that we only need to switch one control animal to positive in the first stratum.

```
input.mtx3 <- input.mtx
input.mtx3[1, 3] <- 1
```

Although the MH estimator still will not work, we can now apply the GN estimator to the full data set:

```
RRstr(Y = input.mtx3)

##
## Homogeneity test not possible because MLE = 1
##
## PF
## 95% interval estimates
##
##          PF    LL UL
## starting   1 0.000  1
## mle        1 0.255  1
## skew corr  1 0.295  1
```

The point estimate is still unity, but the confidence interval is now noticeably narrower than when we threw out the first stratum. This illustrates how sensitive the estimates can be to the endpoint of interest, mortality vs.~morbidity, or put another way, relabeling a single animal from negative to positive.

3.6 A large sample example

This example is from the literature, Stokes et al. (2000,p. ~45). A treatment for respiratory illness was evaluated at two centers. The endpoint is improvement (yes or no). The interest was in estimating the risk ratio, with $RR > 1$ being favorable to the experimental treatment. Here are the data:

```
input.mtx <- matrix(c(29, 45, 14, 45, 37, 45, 24, 45), 2, 4, byrow = TRUE)
```

A tabulation of the data (each row is a stratum) is in Table 9.

	Positive vaccinates	Total vaccinates	Positive controls	Total controls
1	29	45	14	45
2	37	45	24	45

Table 9: A large sample example

Here are the risk ratios for individual strata:

```
rr.by.stratum <- (input.mtx[, 1] / input.mtx[, 2]) /
  (input.mtx[, 3] / input.mtx[, 4])
round(rr.by.stratum, 2)
```

```
## [1] 2.07 1.54
```

There are only two strata, not enough to adequately assess homogeneity. The Mantel-Haenszel estimate is

```
RRmh(Y = input.mtx, pf = FALSE)
```

```
##
## RR
## 95% interval estimates
##
##   RR   LL   UL
## 1.74 1.33 2.27
```

This matches the output of SAS Proc Freq (Stokes et al. 2000,p. ~50). Here is the Gart-Nam estimate:

```
RRstr(Y = input.mtx, pf = FALSE)
```

```
##
## Test of homogeneity across clusters
##
## stat   1.04
## df      1
## p       0.309
##
## RR
## 95% interval estimates
##
##           RR   LL   UL
## starting  1.68 1.47 1.92
## mle       1.69 1.45 2.25
## skew corr 1.69 1.51 2.26
```

Let's compare the MH and GN estimates to the pooled estimate. The calculation takes a few moments, so let's time it:

```
system.time(b <- RRtosst(apply(input.mtx, 2, sum), pf = FALSE))
```

```
##   user  system elapsed
## 22.33   0.07   22.75
```

```
b
```

```
##
## RR
## 95% interval estimates
##
##   RR   LL   UL
## 1.74 1.33 2.33
```

The Mantel-Haenszel and pooled estimates match, but the pooled estimate has a slightly wider (most conservative) confidence interval. The Gart-Nam estimate has a lower point

estimate and has the least conservative (narrowest) confidence interval. All of the point estimates are *in between* the individual strata risk ratios, 2.07 and 1.54.

4 Future work

This vignette should be expanded in the future with the following enhancements:

- Comparison of input data formats for `RRmh()` and `mhgr()` (from the `Hmisc` package).
- Comparison of these methods with logistic regression based methods, including dispersion models, for the same example data sets.

Code in this vignette was run using version 9.5.4 of the PF package available via [github](#).

References

- Agresti, A. (2002), *Categorical data analysis*, Hoboken: Wiley.
- Agresti, A., and Hartzel, J. (2000), “Strategies for comparing treatments on a binary response with multi-centre data,” *Statistics in Medicine*, 19, 1115–1139.
- Carter, R., Lipsitz, S., and Tilley, B. (2005), “Quasi-likelihood estimation for relative risk regression models,” *Biostatistics*, 6, 39–44.
- Gart, J. (1962), “On the combination of relative risks,” *Biometrics*, 18, 601–610.
- Gart, J. (1985), “Approximate tests and interval estimation of the common relative risk in the combination of 2×2 tables,” *Biometrika*, 72, 673–677.
- Gart, J., and Nam, J.-M. (1988), “Approximate interval estimation of the ratio of binomial parameters: A review and corrections for skewness,” *Biometrics*, 44, 323–338.
- Greenland, S., and Robins, J. (1985), “Estimation of a common effect parameter from sparse follow-up data,” *Biometrics*, 41, 55–68.
- Heinze, G., and Schemper, M. (2002), “A solution to the problem of separation in logistic regression,” *Statistics in Medicine*, 21, 2409–2419.
- Hirji, K. (2006), *Exact analysis of discrete data*, Boca Raton: Chapman & Hall/CRC.
- Kuritz, S., Landis, J., and Koch, G. (1988), “A general overview of mantel-haenszel methods: Applications and recent developments,” *Annual Reviews of Public Health*, 9, 123–160.
- Lachin, J. (2011), *Biostatistical methods: The assessment of relative risks*, Hoboken: Wiley.
- Mantel, N., and Haenszel, W. (1959), “Statistical aspects of the analysis of data from

- retrospective studies of disease,” *Journal of the National Cancer Institute*, 22, 719–748.
- McCullagh, P., and Nelder, J. (1989), *Generalized linear models*, London: Chapman; Hall.
- McCulloch, C., Searle, S., and Neuhaus, J. (2008), *Generalized, linear, and mixed models*, Hoboken: Wiley.
- Radhakrishna, S. (1965), “Combination of results from several 2×2 contingency tables,” *Biometrics*, 21, 86–98.
- Stokes, M., Davis, C., and Koch, G. (2000), *Categorical data analysis using the sas system*, Cary, NC: SAS Institute.
- Wacholder, S. (1986), “Binomial regression in glim: Estimating risk ratios and risk differences,” *American Journal of Epidemiology*, 123, 174–184.
- Wedderburn, R. (1974), “Quasilikelihood functions, generalized linear models, and the gauss-newton method,” *Biometrika*, 63, 27–32.
- Yelland, L., Salter, A., and Ryan, P. (2011), “Relative risk estimation in randomized controlled trials: A comparison of methods for independent observations,” *International Journal of Biostatistics*, 1, 5.
- Zou, G. (2004), “A modified poisson regression approach to prospective studies with binary data,” *American Journal of Epidemiology*, 159, 702–706.