The objective is to estimate the average prevalence of correlated binary outcomes. Specifically, assuming that there are patients and there are exchangeable binary outcomes per patient denoted by the parameter of interest is . We want to estimate this parameter and construct its 95% confidence interval. The observed data can be summarized by where is the number of patients with exactly positive responses. Under the exchangeability assumption:do we actually need the exchangeability assumption? The vectors are IID, so the vector sums are IID in {0,1,2,3}, which can be viewed as categories of a multinomial.

where Under this model, the parameter of interest

To this end, we first consider a test statistic for testing the null hypothesis

where

The exact p-value can be calculated by

where the probability is with respect to the random variable

where

and

If is available, then we can reject or accept the hypothesis that depending on if

To calculate we need to calculate for all such that which is infeasible in practice. Instead, we can select “dense” grid sets to cover the parameter space

For a given we can select a small number and approximate by

where Here can be easily calculated by a Monte-Carlo method. Specifically, we can simulate a large number of ), and calculate the corresponding is the proportion of s greater than observed

The 95% confidence interval of is then can be constructed as all with an estimated

The entire procedure can be described in the following algorithm

* For
  + Simulate from unit exponential distribution and let
  + For
    - Simulate
    - Calculate
    - Calculate
    - Calculate where does the variance formula come from? seems that \tilde{p} gives the multinomial covariance matrix. But this doesn’t seem to match eg binomial terms on the diagonal. Maybe cancellations?
* For
  + Identify all such that
  + For each identified 
    - Calculate Shouldn’t this theta refers to theta computed form ? is sampled under the theta computed fom **,** so that is the theta that should be used for centering under the null, right? And not the theta in “For
    - Calculate the proportion denoted by
  + Calculate
* Denote the resulting p values by and the final 95% confidence interval for can be constructed as

In addition, we may consider a different test statistic. In particular, when is expected to be very close to 0 or 1, then it can be more appropriate to consider a test statistic in the form of