A SAS IML Macro for Performing Koch's Stratified Test of Equivalence

Jimmy Thomas Efird
Stanford School of Medicine

Introduction

In this paper, we describe a method for evaluating incidence rates amongst a stratified reference and suspect population with the goal of establishing the statistical equivalence between the groups. Unlike traditional surveillance or intervention methods which focus on detecting a difference in rates, the underlying premise of the equivalence method is to establish that the incidence rates between populations are the same within a predefined range of equivalence, e.g., $-\Delta$ to $+\Delta$. Failure to detect a difference using a comparative means does not infer equivalence, but may merely reflect a lack of power to show that a difference exists. On the other hand, absolute equivalence is impossible to demonstrate, in that equivalence is defined relative to stated error probabilities and the resultant size of the groups. Deviations in conduct that effectively reduce variation may lead to false conclusions of equivalency. A finding of equivalence does not imply lack of effect, as both groups may be equally effective.

Scope of Equivalence Method

The equivalency framework is especially useful in two contexts for the health care analyst. The first is that of an intervention study, wherein an alternative course of action is deemed equally effective to a reference standard in containing a medical condition, but is less costly to administer or lacks negative effect. Under the second scenario, a suspect population is monitored with respect to a reference standard, with no intention of modifying the condition or implementing risk factor analysis if the groups are shown to be statistically equivalent.

Methodology

Two cases of the equivalence hypothesis exists, depending upon whether the construct is partitioned into an inferiority, equivalence, and superiority decision space, in contrast to indicating inferiority to the standard.

Letting P_r and P_s denote the event rates in the reference and suspect groups, and assuming that the difference between P_r and P_s is normally distributed, a decision theoretic hypothesis in the first case is given as:

$D_o : P_r - P_s < -\Delta$ or $P_s > P_r + \Delta$ (Suspect is inferior to reference standard)

$D_1 : -\Delta \le P_r - P_s \le +\Delta$ (Suspect is equivalent to reference standard)

$D_2 : P_r - P_s + \Delta$ or $P_s < P_r - \Delta$ (Suspect is superior to reference standard)

A single decision regarding the difference in rates is obtained based on the $100(1-2\alpha)\%$ confidence interval for $P_r - P_s$. Inequivalence is established if either the upper limit is less than $-\Delta$ (accept D_o) or the lower limit is greater than $+\Delta$ (accept D_2). When the confidence interval is within the

range $[-\Delta \text{ to } +\Delta]$, the groups are deemed equivalent (accept D_1). The result is equivocal if the confidence interval extends beyond a single acceptance region.

Given (j) stratum and letting P_{jr} and P_{js} denote the event rates in stratum j from Groups R and S, the $100(1-2\alpha)\%$ confidence interval is computed as:

$$[\Sigma W_j (P_{jr} - P_{js}) / \Sigma W_j] \pm Z_{(1-\alpha)}\sqrt{V}$$

where

$$W_j = (n_{jr} * n_{js})/(n_{jr} + n_{js})$$

$$V = \Sigma W_j^2 V_j / \Sigma (W_j)^2$$

$$V_j = [P_{jr} (100 - P_{jr})/n_{jr}] + [P_{js} (100 - P_{js})/n_{js}]$$

and

$$\alpha = 1 - \text{Prob}(X \le -Z_{(1-\alpha)}), \quad X \approx N(0,1).$$

When there is only one stratum and the suspect and reference standard groups are of equal size, the above formula simplifies to:

$$P_r - P_s + Z_{(1-\alpha)}\sqrt{[P(100-P)]}.$$

To show that the suspect group is not inferior to the reference standard group, the hypothesis is stated in traditional terms as:

$H_o : P_r - P_s > \alpha$ (Inequivalence)

$H_1 : P_r - P_s < \alpha$ (Equivalence)

with the corresponding 1-sided 100 (1 - □) % confidence interval calculated as above.

Example

A public health official has identified 15 counties throughout the state as having unacceptably high rates of a contagious disease. The official is interested in conducting an intervention study to determine if a directed letter with educational material outlining the detrimental aspects of the disease, will substantially reduce the proportion of cases. Two similar health care providers, each servicing the suspect counties, were selected for the study. One of the providers served as a control, while intervention techniques were applied to the other provider. The data for this study is shown in Figure 1. The SAS code for calculating the 2-sided 95% confidence interval is given in Figure 2. Examining the results, we see that an equivalent proportion of cases was detected between the two providers, as the 95% confidence interval, e.g., [-8.32732, 7.1961776] was fully contained within the equivalence limits of [-10, +10], e.g., D_1 accepted.

Figure 1: Data Results

County	Group	Size	Number of Confirmed Cases
1	1	12	2
1	2	12	5
2	1	11	0
2	2	10	0
3	1	5	2
3	2	6	1
4	1	23	4
4	2	24	3
5	1	12	2
5	2	16	5
6	1	12	2
6	2	10	2
7	1	25	4
7	2	26	3
8	1	7	2
8	2	8	1
9	1	17	2
9	2	17	4
10	1	25	5
10	2	21	3
11	1	2	0
11	2	2	0
12	1	8	2
12	2	8	2
13	1	12	5
13	2	15	5
14	1	6	2
14	2	6	0
15	1	8	2
15	2	6	4

Figure 2: SAS Code for Stratified Test of Equivalence

```
options nocenter ls=132 mprint symbolgen nodate nonumber cc=cr;
libname d "prd_bt1:[ujte.efird]";
libname library "prd_bt1:[ujte.efird.book]";
data a;
   infile 'prd_bt1:[ujte.efird.book.chap4]equiv.dat';
   input @1 stratum @5 group @10 event;
proc sort data=a;
   by stratum group;
proc means n mean noprint data=a;
   by stratum group;
      var event;
         output out=out1 n=n mean=m;
proc freq data=a;
   tables stratum*group*event/missprint;
data grp1(drop=group) grp2(drop=group);
   set out1 (keep=group n m);
      m=m*100;
      if          group=1 then output grp1;
         else if group=2 then output grp2;
proc iml worksize=500;
   start main;
      alpha=0.025;
      use grp1;
      read all into grp1;
      use grp2;
      read all into grp2;
      wj=(grp1[,1]#grp2[,1])/(grp1[,1]+grp2[,1]);
      vj=((grp1[,2]#(100-grp1[,2]))/grp1[,1])+((grp2[,2]#(100-grp2[,2]))/grp2[,1]);
      pdiff=grp1[,2]-grp2[,2];
      dmean=(wj`*pdiff)/wj[+];
      v=((wj##2)`*vj)/(wj[+]##2);
      lcl=dmean-probit(1-alpha)#sqrt(v);
      ucl=dmean+probit(1-alpha)#sqrt(v);
      print lcl;
      print ucl;
   finish;
   run main;
```

Summary

The decision theoretic approach to equivalence testing offers an appealing means for partitioning the decision space in a region allowing for inferiority, equivalence, and superiority. Although further study is needed of closed sequential methods to address sample size inflation issues, this approach should perform reasonably well in most practical health care applications.

References

Blackwelder W. Proving the null hypothesis in clinical trials. Controlled Clinical Trials. 1982;3:345-53.

Fleiss J. Statistical methods for rates and proportions. New York. John Wiley and Sons, Second Edition. 1989.

Jones B, Jarvis P, Lewis J, and Ebbutt A. Trials to assess equivalence: The importance of rigorous methods. BMJ 1996;313:36-39.

Kock GG, Amara IA, Stokes ME, and Uryniak TJ. Categorical data analyses, statistical methodology in the pharmaceutical sciences. Edited by Donald A. Berry, pages 389-473. Marcel Dekker Inc., New York, 1990.