PharmaSUG 2019 - Paper ST-314

Application of Criterion I^2 in Clinical Trials Using SAS®

Igor Goldfarb, Accenture ALSS;
Mitchell Kotler, PRA

ABSTRACT

The goal of this paper is to demonstrate how meta-analyses criteria (specifically, I^2 and Cochran's Q) can be calculated using SAS® software and applied to the examination of the heterogeneity across subgroups. Some European regulatory authorities require an analysis of the subgroups effect estimated using the criterion I^2, which was developed and suggested by Higgins and his colleagues to provide a researcher with a better measure of the consistency between trials in a meta-analysis. Originally the criterion I^2 was developed to estimate heterogeneity across the studies, whereas the authorities required its application to the scrutiny across subgroups of the same study. The criterion I^2 describes the percentage of total variation across studies that is due to heterogeneity rather than chance. Negative values of I^2 are put equal to zero so that I^2 lies between 0% and 100%. A value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity. SAS® programs were developed to implement the algorithm to calculate I^2 and similar parameters (e.g., Cochran's Q). An analysis showed that the SAS® code successfully reproduced published results that were produced using other statistical software. After verification the SAS® code was run on the clinical data and the results obtained were successfully submitted to authority.

INTRODUCTION

When researchers are discussing any important effect (e.g., treatment effect on efficacy endpoint), an assessment of the consistency of this effect across multiple studies (normally made by various investigators under different conditions) is one of the vital things the scientific community would like to know. Evaluation of a homogeneity/heterogeneity (the term heterogeneity will be further used in this paper) represents an essential part of meta-analysis of the given set of studies under consideration. Recently, meta-analyses and its modifications attract more attention of researchers working in the pharmaceutical industry and specialists from regulatory agencies. In some cases, a traditional understanding of meta-analysis was transformed and interpreted more widely (or, in some sense, narrower). Specifically, the same methodology of investigating heterogeneity across different studies was applied to heterogeneity across subgroups within one clinical trial. Very extensive research on meta-analysis in application to subgroups can be found in a publication by Borenstein and Higgins (2013). Some European regulatory authorities require this type of analysis of heterogeneity as a part of the submission package. Here we present the SAS® code (developed by the authors according to the published algorithms – see Appendix) for two statistical tests that are commonly used for meta-analysis (Cochran Q and I^2).

TESTS FOR HETEROGENEITY

A number of different tests for heterogeneity have been proposed. For example, Higgins *et al.,* 2002, suggests three different statistic for this purpose – H^2, R^2 and I^2, all in addition to existing criterion – Cochran Q. The authors of the present paper are concentrated on the two statistical tests that at present are commonly used for meta-analysis in the pharmaceutical research - Cochran Q and *I*-squared (I^2). As any other standard statistical approach, the test for heterogeneity examines the null hypothesis that all studies are evaluating the same effect. The classical test statistic (Cochran's Q) is lately accompanied by I^2 criterion that was suggested by Higgins and his colleagues (2002, 2003).

I^2 TEST

As it was mentioned above Higgins and his colleagues (2002, 2003) developed and suggested a few statistical approaches that quantify the heterogeneity of the given set of studies. We focus on one of them – I^2. The criterion I^2 can be readily calculated from basic results obtained from a typical meta-analysis as

$$I^2 = \frac{Q - df}{Q} \times 100\%$$

where Q is Cochran's heterogeneity statistic and *df* the degrees of freedom. Negative values of I^2 are forcefully set equal to zero so that numerical values of the criterion I^2 always lies between 0% and 100%. The calculated quantity describes the percentage of total variation [of the treatment effect] across studies that is due to heterogeneity rather than chance. Therefore, a value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity. One can accept a naïve categorization of values for I^2 as low (25%), moderate (50%) and high (75%) as suggested by Higgins *et al*, 2003, but it is evident that this classification would not be appropriate for all situations and accompanied circumstances.

One of the main advantages of the criterion I^2 is that its values can be directly compared between meta-analyses with different number of studies and different types of outcome data.

CALCULATION AND VERIFICATION

According to the aforementioned formula for the criterion I^2 its calculation is based on Cochran Q value. Therefore, a SAS® program was developed to obtain this quantity (and, correspondingly, I^2) based on the well-known algorithm, because Cochran Q is not directly produced by any of the standard statistical procedures in SAS®. Cochran's Q is computed as a weighted sum of the squared deviations of each study's estimate from the overall meta-analytic estimate, weighting each study's contribution in the same manner as in the meta-analysis.

To verify the developed program we apply the code to the analysis of the example that was considered by Higgins et al (2003) - meta-analysis of randomized controlled trials of amantadine for preventing influenza. The original publication was published by Jefferson et al. in 2002. Table 1 displays title of the studies involved and appropriate outcomes

Using the results from the 8 studies shown in Table 1 below, we computed Cochran Q = 12.2, and corresponding I^2 = 42.6% for 7 df. The value Q = 12.4 with I^2 = 43.5% was published in Higgins (2003), based on the Mantel-Haenszel estimate of the fixed-effect meta-analysis (rather than the usual inverse variance weighted average); this can also be obtained by adjusting the weighting in our code.

Trial	Drug (n/N)	Placebo (n/N)
Oker-Blom (1970)	16/141	41/152
Muldoon (1976)	1/53	8/52
Monto (1979)	8/136	28/139
Kantor (1980)	9/59	9/51
Pettersson (1980)	32/95	59/97
Quarles (1981)	15/107	20/99
Dolin (1982)	2/113	27/132
Reuman (1989)	3/317	5/159

Table 1. Eight trials of amantadine for prevention of influenza. Outcome is cases of influenza.

APPLICATION TO SUBGROUP ANALYSIS

When applying the criterion I^2 to the case of subgroups within a single study, the number of degrees of freedom (*df*) is normally small, for example, two gender subgroups- males and females. In its turn it means that for all studies with Cochran Q (calculated for the subgroups within the clinical trial under investigation) greater than 4 computed values of the criterion I^2 will be larger than 75% (remind, this interval of quantities is interpreted as high heterogeneity).

To illustrate this, consider as an example a hypothetical study where 200 males and 200 females were randomized equally to the experimental drug and placebo. Successes and failures are as shown below in Table 2.

Subgroup	Drug (n/N)	Placebo (n/N)
Females	50/100	40/100
Males	40/100	50/100

Table 2. Sample study 1 (hypothetical) – two subgroups.

The calculated result for Cochran Q obtained on the hypothetical data presented in the Table 2 is 4.0 and I^2=75% (*df*=1). The conclusion about high heterogeneity has to follow, but intuitive glance at the data does not witness that the distribution is very heterogeneous.

The authors believe that this contradiction goes well with the known fact that Cochran Q test is known to be poor at detecting true heterogeneity as significant. The power of the test in our case is low due to a number of subgroup (similar to regular meta-analysis with very small number of studies).

Despite all reservations and potential pitfalls (only small part of them is discussed in this chapter) the developed program was used to address the request from European regulators. The code (after appropriate verification and validation) was successfully applied to the actual analysis of the heterogeneity of the efficacy and safety results across standard (gender, race, age, etc.) and disease-related subgroups for a number of individual clinical trials and the outputs were submitted to the authority.

CONCLUSIONS

To recap the authors would like to underline the following:

1. The suggested earlier meta-analyses criteria (specifically, I^2 and Cochran's Q) can be calculated using SAS® software – a program was developed.
2. The developed code can be applied to the examination of the heterogeneity across individual studies. Available (published) actual data were used to verify the program and the calculated results coincide with outcomes that were independently obtained by other authors.
3. An application to subgroup analysis was initially investigated using hypothetical set of clinical data. These results should be interpreted with caution in cases of a small number of subgroups.
4. The developed code was applied to the actual analysis of the heterogeneity of the results across standard and disease-related subgroups for data collected in actual clinical trials and the outputs were successfully submitted to the authority.

REFERENCES

Borenstein, M., and Higgins, J.P.T. (2013), “Meta-analysis and subgroups”, Prev Sci DOI 10.1007/s11121-013-0377-7.

Higgins, J.P.T., and Thompson, S. (2002), “Quantifying heterogeneity in a meta-analysis”. *Statistics in medicine*. 21, 1539–1558.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., and Altman, D.G. (2003), “Measuring inconsistency in meta-analysis”. BMJ 327, 557-560.

Jefferson, T.O., Demicheli, V., Deeks, J.J., Rivetti, D. (2002) “Amantadine and rimantadine” for preventing and treating influenza A in adults”. *Cochrane Database Syst Rev* 4: CD001169.

ACKNOWLEDGMENTS

The authors are very thankful to Prof Julian Higgins for the fruitful discussion of the applicability of heterogeneity criteria to the investigation of the subgroup effects. The authors are also very thankful to Jack Jacobsen, Senior Director of Clinical Programming and Statistics (Accenture ALSS), for this constant support of this work.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please feel free to contact anyone of the authors at:

Igor Goldfarb
Accenture ALSS
igor.goldfarb@accenture.com

Mitchell Kotler
PRA
kotlermitch@prahs.com

APPENDIX

Here is the program that was used to calculate a value of Cochran Q using the data under consideration (note that this variant of the code uses usual inverse-variance weighted average instead of Mantel-Haenszel estimates that was used by Higgins et al (2003)).

```
data indata_a;
  input study_num 1-2 study_name $ 5-14 study_year $ 20-25 D_success 30-32
Total_D 35-37 P_success 40-42 Total_P 45-47;
  trt = 'DRUG';
    do i=1 to D_success; outcome="1-S"; output; end;
    do i=(D_success+1) to Total_D; outcome="2-F"; output; end;
  trt = 'PLAC';
    do i=1 to P_success; outcome="1-S"; output; end;
    do i=(P_success+1) to Total_P; outcome="2-F"; output; end;
  *drop i D_success Total_D P_success Total_P;
*        1         2         3         4         5;
*234567890123456789012345678901234567890123456789012345678901234567890;
```

```
datalines;
01  Oker-Blom      (1970)    016  141  041  152
02  Muldoon        (1976)    001  053  008  052
03  Monto          (1979)    008  136  028  139
04  Kantor         (1980)    009  059  009  051
05  Pettersson     (1980)    032  095  059  097
06  Quarles        (1981)    015  107  020  099
07  Dolin          (1982)    002  113  027  132
08  Reuman         (1989)    003  317  005  159
;
run;
data indata _b;
  set indata_a;
  drop i D_success Total_D P_success Total_P;
run;

proc freq data=indata _b;
  by study_num;
  tables trt*outcome / nopercent nocol cmh;
  *output out=pooled all;
run;

proc freq data=indata _b;
  by study_num;
  tables trt*outcome / nopercent nocol cmh;
  *output out=pooled all;
  ods output CMH=c_m_h CommonRelRisks=c_r_r;
run;

proc freq data=indata _b;
  *by study_num;
  tables trt*outcome / nopercent nocol cmh;
  *output out=pooled all;
  ods output CMH=c_m_h_tot CommonRelRisks=c_r_r_tot;
run;

data c_r_r_a;
  set c_r_r(where=(statistic="Odds Ratio"));
run;

data logodds;
  set c_r_r_a;
  log_odds = log(value);
  log_lower = log(lowercl);
  log_upper = log(uppercl);
  variance = ((log_upper - log_odds)/1.96)**2;
  wgt = 1/variance;
  weighted = log_odds*wgt;
  keep study_num log_odds log_lower log_upper variance wgt weighted;
run;

proc means data=logodds;
  var wgt weighted log_odds;
  output out=sums sum=sum_wgt sum_weighted sum_log_odds;
run;
```

```
data deviations;
  merge sums(in=use keep=sum_wgt sum_weighted sum_log_odds) logodds;
  if use then w_mean = sum_weighted/sum_wgt;
  w_mean_or = exp(w_mean);
  log_odds_mean = exp(sum_log_odds);
  dev = log_odds - w_mean;
  chi_cont = dev**2/variance;
  retain w_mean;
run;

proc means data=deviations;
  var chi_cont;
  output out=Q_value sum=Q;
proc print data=Q_value;
  var Q;
run;
```