

Common and Comparative Incidence Indicators of Adverse Events for Well-Defined Study Pools

John R Gerlach, Dataceutics, Inc., Pottstown, PA USA

ABSTRACT

Consider a large Integrated Safety Summary (ISS) study where the analysis of adverse events (AE) is based on well-defined study pools, such that each pool represents a unique collection of subjects. A subject might belong to more than one study pool, perhaps having a particular indication or some other criterion of clinical interest. Besides the standard adverse event reports, that is, a categorized event across treatment groups, it may be required to show incidence with respect to the degree of incidence, very common versus common, as well as in comparison between the study drug and placebo. The task of imputing so-called Common and Comparative incidence indicators becomes more involved since each indicator is determined by its own study pool. Moreover, the workload and its complexity have increased substantially. This paper explains the process of expanding the data set of adverse events (ADAE) and imputing these indicator variables for subsequent analysis.

INTRODUCTION

Imagine a large ISS study where the analyses are based on study pools that represent sets of subjects having specific attributes of clinical interest. For example, a study pool might identify subjects belonging to several component core studies or, perhaps, a specific indication, such as COPD or epilepsy. Conversely, the study pool might exclude certain subjects, such as those subjects who took the study drug in the core study but not in the follow-up study. Whatever the criteria, the study pools are *well-defined*. Moreover, the study pool might not be mutually exclusive; that is, a subject can be found in more than one study pool. Consequently, a single study might contribute to more than one study pool, such as the subjects in Study ABC001 contributing to Study Pools SP01 and SP03.

Having three study pools implies three analysis sets, which triples the number of reports. For example, there would be three versions of the standard report of Adverse events by MedDRA categories across treatment groups, as indicated below by the subtitle “*Analysis Set: Study Pool S1.*” Besides the additional standards reports, however, there are additional reports (in triplet) pertaining to the Common and Comparative indicators. In fact, there may be a report that involves both Common and Comparative indicators, perhaps having the title, “*Incidence of Common and Very Common TEAEs with 1% or Greater Incidence Compared to Placebo.*” Obviously, the complexity and amount of work has increased.

Table 4.3.1
Incidence of TEAEs by Incidence Categories
Analysis Set: Study Pool S1

MedDRA (Version 15.0) Primary System Organ Class Preferred Term	Placebo N=XXX n (%)	ABC 50mg N=XXX n (%)	ABC 200mg N=XXX n (%)	ABC Overall N=XXX n (%)
SYSTEM ORGAN CLASS 1	XXX (XX.X)	XXX (XX.X)	XXX (XX.X)	XXX (XX.X)
PREFERRED TERM 1	XXX (XX.X)	XXX (XX.X)	XXX (XX.X)	XXX (XX.X)
PREFERRED TERM 2	XXX (XX.X)	XXX (XX.X)	XXX (XX.X)	XXX (XX.X)
:	:	:	:	:

THE STUDY POOLS

The study pools are identified by the variables PSnFL and PSnFN, containing the values ‘Y’ or ‘N’ and their numeric counterparts, 1 and 0, respectively. Thus, for example, the variables PS1FL identifies those subjects belonging to Study Pool 1 having the value ‘Y’; whereas, the variable PS2FL indicates Study Pool 2, accordingly. These variables should be stored in the ADaM ADSL (Subject Level) data set, since they denote subject populations. Consider the following distribution of these variables juxtaposed with the Study Identifier. Notice that Study Pools 1 and 2 are mutually exclusive; however, Study Pool 3 represents subjects from the first two pools.

STUDYID	PS1FL	PS2FL	PS3fl	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ABC001	N	Y	N	24	0.45	24	0.45
ABC001	N	Y	Y	369	6.92	393	7.37
ABC002	N	Y	N	36	0.68	429	8.05
ABC002	N	Y	Y	441	8.27	870	16.32
ABC003	Y	N	N	30	0.56	900	16.88
ABC003	Y	N	Y	948	17.78	1848	34.67
ABC004	Y	N	N	30	0.56	1878	35.23
ABC004	Y	N	Y	918	17.22	2796	52.45
ABC005	N	Y	Y	1134	21.27	3930	73.72
ABC006	N	Y	Y	207	3.88	4137	77.60
ABC007	Y	N	N	81	1.52	4218	79.12
ABC007	Y	N	Y	1113	20.88	5331	100.00

Table 1. Study pools.

BUILDING THE INCIDENCE DATA SET

Expanding the ADAE data set requires little more than a Data step, as shown below. The MERGE statement joins the AE data and the study pool variables from ADSL. The sub-setting-IF statement ensures that only those subjects who had an AE are processed. Then the PS n FL variables are used to output observations, accordingly. Notice that the target data set INCIDENCE has only the variables POOL, denoting the study pool, along with three other relevant variables denoting the event category (Body System / Preferred Term) and whether the AE is treatment emergent.

```

data incidence01;
  merge adae(keep=usubjid aebodsys aedecod trtemf in=ae)
        adsl(keep=usubjid ps1fl ps2fl ps3fl);
  by usubjid;
  if ae;
  if ps1fl eq 'Y' then do; pool = 'PS1'; output; end;
  if ps2fl eq 'Y' then do; pool = 'PS2'; output; end;
  if ps3fl eq 'Y' then do; pool = 'PS3'; output; end;
  keep pool aebodsys aedecod trtemf;
run;

```

Once the incidence data set has been created, the next two steps compute the incidence (in percentages) of the events, categorized by body system / preferred term, for each study pool. The percentage is used to determine the levels of incidence: Very Common, Common, and Uncommon. Notice that the denominator is the number of AEs in the study pool, not the number of subjects in the study pool.

```

proc freq data=incidence01 noprint;
  tables pool / out=tots(drop=percent rename=(count=total));
  tables pool * aebodsys * aedecod / out=cnts(drop=percent);
  where trtemf eq 'Y';
run;

data pct;
  merge cnts tots;
  by pool;
  pct = (count / total) * 100;
run;

```

IMPUTING COMMON INCIDENCE INDICATORS

The Common incidence indicator variables are assigned based on the following cut-off values prior to rounding of the percentages, albeit with adequate decimal places for reporting to avoid misinterpretation.

- Very Common Incidence rate of at least 10%
- Common Incidence rate of at least 1%, but less than 10%
- Uncommon Incidence rate less than 1%

There are three study pools, hence three indicator variables, PS n COMM, where n denotes the study pool. The data set PCTS contains a record for each categorized event manifested within a given study pool, along with the respective percentage of incidence. Keep in mind that the data set is *not* subject level; rather, the unit of analysis is the *study pool / adverse event*. Thus, for a given observation, it is necessary to determine the study pool, for which the SELECT/WHEN statement is very apropos, then the cutoff logic is applied, albeit assigning the appropriate indicator variable (e.g. PS1COMM), accordingly. Notice that the Data step creates three output data sets, one for each study pool.

```
data ps1(keep=aebodsys aeecod ps1comm)
  ps2(keep=aebodsys aeecod ps2comm)
  ps3(keep=aebodsys aeecod ps3comm);
length ps1comm ps2comm ps3comm $25;
set pcts;
select(pool);
  when('PS1') do;
    if pct ge 10
      then ps1comm = 'VERY COMMON';
    else if 1 le pct lt 10
      then ps1comm = 'COMMON';
    else ps1comm = 'UNCOMMON';
    output ps1;
  end;
  when('PS2') do;
    if pct ge 10
      then ps2comm = 'VERY COMMON';
    else if 1 le pct lt 10
      then ps2comm = 'COMMON';
    else ps2comm = 'UNCOMMON';
    output ps2;
  end;
  when('PS3') do;
    if pct ge 10
      then ps3comm = 'VERY COMMON';
    else if 1 le pct lt 10
      then ps3comm = 'COMMON';
    else ps3comm = 'UNCOMMON';
    output ps3;
  end;
  otherwise;
end;
label ps1comm = 'Common Incidence Indicator-Pool S1'
      ps2comm = 'Common Incidence Indicator-Pool S2'
      ps3comm = 'Common Incidence Indicator-Pool S3';
run;
```

The Data step below merges the three data sets PS1, PS2, and PS3 so that the data PSCOMM represents all the adverse events, along with the three Common indicator variables. Keep in mind that the data set indicates the *categorical level* of incidence for each event, per study pool. In fact, it is conceivable that an adverse event did not manifest for a specific study pool; consequently, the indicator variable would contain a null value. The next Data step merges these indicator variables with the AE data set, by System Organ Class / Preferred Term. However, the task is not yet complete. Why? Because these variables must be assigned based on whether the AE is treatment emergent, *which is germane to a subject*, not the event.

```

data pscomm;
  merge ps1 ps2 ps3;
  by aebodsys aeecod;
run;

data adae02;
  merge adae01(in=ae) pscomm;
  by aebodsys aeecod;
  if ae;
run;

```

The next Data step assigns the Common indicator variables properly, that is, based on whether the subject's adverse event is treatment emergent. Notice that if the event (for a given subject) is not treatment emergent then the PS_nCOMM variable is appropriately re-assigned a null value *for that subject*. This explains the need to separate the tasks of merging the event-based Common indicator variable and, then, assigning the variables to null, as needed. Also, the Data step creates the respective numeric version, PS_nCOMM_N.

```

data adae03;
  set adae02;
  if trtemfl eq 'Y' and ps1fl eq 'Y'
  then do;
    if ps1comm eq 'VERY COMMON'
    then ps1commn = 1;
    else if ps1comm eq 'COMMON'
    then ps1commn = 2;
    else if ps1comm eq 'UNCOMMON'
    then ps1commn = 3;
  end;
  else ps1comm = '';
  if trtemfl eq 'Y' and ps2fl eq 'Y'
  then do;
    if ps2comm eq 'VERY COMMON'
    then ps2commn = 1;
    else if ps2comm eq 'COMMON'
    then ps2commn = 2;
    else if ps2comm eq 'UNCOMMON'
    then ps2commn = 3;
  end;
  else ps2comm = '';
  if trtemfl eq 'Y' and ps3fl eq 'Y'
  then do;
    if ps3comm eq 'VERY COMMON'
    then ps3commn = 1;
    else if ps3comm eq 'COMMON'
    then ps3commn = 2;
    else if ps3comm eq 'UNCOMMON'
    then ps3commn = 3;
  end;
  else ps3comm = '';
  label ps1commn = 'Common Incidence Indicator - Pool S1 - num'
        ps2commn = 'Common Incidence Indicator - Pool S2 - num'
        ps3commn = 'Common Incidence Indicator - Pool S3 - num';
run;

```

The cardinality of the ADAE data set is unchanged, however, there are new indicator variables denoting the categorical level of incidence for treatment emergent AEs that are specific to a given study pool. Thus, these indicator variables should be null for non-treatment emergent AEs.

The next report shell illustrates the purpose of creating the PS_nCOMM variables. Notice that the report is similar to the standard report showing the incidence AEs by MedDRA category across treatment groups. However, the events are listed in the context of their categorical level of incidence. From a clinical perspective, the report distinguishes events within a body system, emphasizing the level of incidence. Indeed, this report emulates the standard report for severity (i.e. Mild, Moderate, and Severe), for example.

Table 4.3.2
Incidence of TEAEs by Incidence Categories
Analysis Set: Study Pool S1

MedDRA (Version 15.0) Primary System Organ Class Preferred Term	Placebo N=XXX		ABC 50mg N=XXX		ABC 200mg N=XXX		ABC Overall N=XXX	
	n	(%)	n	(%)	n	(%)	n	(%)
SYSTEM ORGAN CLASS 1	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)
VERY COMMON								
PREFERRED TERM 1	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)
PREFERRED TERM 2	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)
:	:	:	:	:	:	:	:	:
COMMON								
PREFERRED TERM 1	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)
PREFERRED TERM 2	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)
:	:	:	:	:	:	:	:	:
UNCOMMON								
PREFERRED TERM 1	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)
:	:	:	:	:	:	:	:	:

COMPARATIVE INDICATORS

Similar to the Common indicator variables that indicated a ratio as percentages, the Comparative indicator variables represent the percentage difference between study drug and placebo, that is, a dichotomous variable indicating a threshold of 1% or greater. Unlike the Common variables, the cutoff will be determined after the rounding of percentages, rounded to the nearest integer. As shown below, the report will be similar to the standard report listing categorized AEs across treatment groups; however, it will represent only those events where the incidence for the study drug was at least 1% or greater than Placebo, for each Study Pool (note the subtitle of the report). Thus the initial task is to create these indicator variables.

Table 4.3.3
Incidence of TEAEs with Incidence of **at Least 1% Greater Than Placebo**
Analysis Set: Study Pool S1

MedDRA (Version 15.0) Primary System Organ Class Preferred Term	Placebo N=XXX		ABC 50mg N=XXX		ABC 200mg N=XXX		ABC Overall N=XXX	
	n	(%)	n	(%)	n	(%)	n	(%)
SYSTEM ORGAN CLASS 1	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)
PREFERRED TERM 1	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)
PREFERRED TERM 2	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)	XXX	(XX.X)

First it is necessary to subset the incidence data set so that it represents only treatment emergent events. Then, *assume for this study* that the incidence data set contains a *date* variable FABCDOSE denoting the first dose date for those subjects who *actually took the study drug* in a core study; whereupon the date variable would be null, otherwise. The subsequent Data step splits the incidence data set accordingly, creating two new data sets ABC (Study Drug) and PBO (Placebo), representing those subjects having had the study drug in the core study and those who took only placebo. Once again, the intermediate data set contains only those relevant variables for processing.

```
proc sort data=incidence01 out=incidence02 nodupkey;
  by usubjid aebodsys aedecod pool;
  where aedecod is not null and trtemfl eq 'Y';
run;

data abc01 pbo01;
  set incidence01;
  if < First dose date for subjects having study drug in core study > not null
    then output abc01;
  else output pbo01;
  keep pool aebodsys aedecod trtemfl;
run;
```

Next, it is necessary to compute the incidence of those subjects who took the study drug and those who took only placebo, as percentages of those subjects in each study pool. Again, the *tots* data sets represent the number of subjects in each study pool. Once the percentages are obtained, this information is merged in order to compute the percentage difference, denoted by the variable PCT_DIFF.

```

proc freq data=abc01 noprint;
  tables pool * aebodsys * aeecod / out=abc02(drop=percent rename=(count=cnt_abc));
run;

data abc03;
  merge tots_abc abc02;
  by pool;
  if cnt_abc ne .
    then pct_abc = (cnt_abc / total) * 100;
run;

proc freq data=pbo01 noprint;
  tables pool * aebodsys * aeecod / out=pbo02(drop=percent rename=(count=cnt_pbo));
run;

data pbo03;
  merge tots_pbo pbo02;
  by pool;
  if cnt_pbo ne .
    then pct_pbo = (cnt_pbo / total) * 100;
run;

data pctdiff01;
  merge abc03 pbo03;
  by pool aebodsys aeecod;
  if pct_pbo eq .
    then pct_pbo = 0;
  if pct_abc ne .
    then do;
      pct_diff = pct_abc - pct_pbo;
      output;
    end;
run;

```

The Data step below easily imputes the comparative indicator variables employing the PCT_DIFF variable. Notice the use of By-group processing at the Body System / Preferred Term level. The data set PCTDIFF02 contains as many as three observations per by-group representing the three study pools. Thus, the target variables are assigned null values at the onset of a by-group, then assigned a 'Y' value if the 1% threshold is met. Consequently, the target data set contains all manifested events along with the three dichotomous variables. Finally, this data set is merged back into the intermediate ADAE data set and is now ready for processing.

```

proc sort data=pctdiff01 out=pctdiff02;
  by aebodsys aeecod;
run;

data pscmp;
  retain ps1cmp ps2cmp ps3cmp ` `;
  set pctdiff02;
  by aebodsys aeecod pool;
  if first.aeecod
    then do;
      ps1cmp = ` `; ps2cmp = ` `; ps3cmp = ` `;
    end;

```

```

select(pool);
  when('PS1')   if pct_diff ge 1 then ps1cmp = 'Y';
  when('PS2')   if pct_diff ge 1 then ps2cmp = 'Y';
  when('PS3')   if pct_diff ge 1 then ps3cmp = 'Y';
  otherwise;
end;
if last.aedecod
  then output;
keep aebodsys aedecod ps1cmp ps2cmp ps3cmp;
label ps1cmp = 'AE Incidence >= 1% of ABC >= PBO Flag - Pool S1'
      ps2cmp = 'AE Incidence >= 1% of ABC >= PBO Flag - Pool S2'
      ps3cmp = 'AE Incidence >= 1% of ABC >= PBO Flag - Pool S3';
run;

data adae04;
  merge adae03(in=ae) pscmp;
  by aebodsys aedecod;
  if ae;
run;

```

CONCLUSION

Common and Comparative indicator variables afford the means for more thorough analysis of adverse events beyond the standard reports of event category across treatment groups. These indicator variables qualify the events prior to the analysis allowing greater granularity or focusing on a subset of events based on a threshold criterion. Also, these variables are study pool specific, thereby expanding analysis even further.

REFERENCES

Analysis Data Model (ADaM) Implementation Guide version 1.0. <http://www.cdisc.org/adam>.

Gerlach, John R; "Imputing Dose Levels for Adverse Events." *Proceedings of the PharmaSUG Conference, 2013*.

Minjoe, Sandra; "Hands-On ADaM ADAE Development." *Proceedings of the PharmaSUG Conference, 2013*.

Minjoe, Sandra, Widel, Mario; "Using the ADaM ADAE Structure for non-AE Data." *Proceedings of the SAS® Global Forum Conference, 2013*.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: John R. Gerlach
E-mail: JRGerlach0907@gmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.