Using Direct Standardization SAS® Macro for a Valid Comparison in Observational Studies

Daojun Mo[1], Xia Li[2] and Alan Zimmermann[1]
[1]Eli Lilly and Company, Indianapolis, IN
[2]inVentiv Clinical Solutions LLC, Indianapolis IN

ABSTRACT

Observational studies are usually imbalanced in the factors associated with the outcome measures. Simply presenting the descriptive results or the P values from an unadjusted between-group comparison could lead to a biased conclusion. Direct standardization is one of the methods for binary data that reveal the valid association between comparison groups. Direct standardization is often implemented in a spreadsheet by copying and pasting the data. This becomes tedious in a study that explores multiple outcome measures. We thus developed a SAS® macro that is adaptable to many types of observational studies which consider binary outcome measures. Examples are given to demonstrate the concept of direct standardization, and how to use the macro.

INTRODUCTION

The essential research question in Phase III clinical trials is to see if a proposed therapy is generally better than the reference one or a placebo or non-inferior to the reference one. Phase IV is important in post-marketing surveillance to monitor the drug safety profile after drug approval, and evaluate effectiveness in actual use. From the study design perspective, Phase IV observational studies usually have less restrictive inclusion/exclusion criteria than a Phase III clinical trial. However, the comparison groups in phase III studies are randomized so that the comparison groups are largely balanced for all factors except the study factor (i.e., therapy). Usually no significant difference can be found in the comparison of the factors other than the study factor. However, Phase IV observational studies along with other observational studies (e.g., medical claim data mining, and public health population study) are not randomized so the comparison groups can be easily imbalanced in the factors that are both associated with the outcome measure and the study groups. These factors, if not taken into consideration, can lead to a biased conclusion. Next an example is given to demonstrate the existence of confounding effects of age and gender, and the use of a standardization method to correct the confounding effects.

PROBLEM EXAMPLE

Hypothetically, a company developed a drug indicated for preventing the development of type 2 diabetes among high-risk adult population (aged ≥20 years) who are obese [body mass index (BMI) >30 kg/m^2]. Regulatory authorities approved the drug for marketing because of the demonstrated efficacy and safety of the drug in pivotal phase III trials. The sponsor company started a phase IV study to make long-term regulatory commitment to safety surveillance on their new drug. The phase IV study enrolled and followed up 3056 patients in a treated group and 818 in an untreated group; all enrolled were obese patients who did not have type 2 diabetes diagnosis at the beginning of the study. The objective was to compare the adverse events profile of the treated obese patients to the untreated obese patients, expecting no difference. In this example, let us focus on the incidence of type 2 diabetes, which the drug is to prevent. One year later, the interim analysis was conducted among those who had sufficient data for clinical evaluation of type 2 diabetes (n=2902) and showed that the raw incidence of type 2 diabetes in the treated group (25.7% (590/2294)) was surprisingly higher than the untreated group (18.4% (112/608)) (P=0.0002 from Chi-square test for a 2x2 table). The higher incidence among treated patients is obviously counterintuitive because the treatment was approved for preventing the occurrence of type 2 diabetes.

A closer look at the risk factors associated with diabetes showed that the patterns of baseline obesity (BMI), physical exercise, and diet are comparable between the treated and untreated groups, but not the factors of age and gender.

> The treated group is older (mean = 54.9, SD=12.4) than the untreated group (43.6, 12.4) (t test P < 0.0001). But age is known to affect type 2 diabetes incidence. And this means that age is both associated with the diabetes incidence and the comparison groups and thus is a confounder of the association between diabetes incidence and treatment. Simply presenting the crude rates of 2 comparison groups could lead to biased overall conclusion.

Similarly, the treated group had more males than the untreated group (66.5% versus 52.3%, chi-square P = 0.0002). But gender is known to affect type 2 diabetes incidence. And this means that gender is both associated with the diabetes incidence and the comparison groups and thus is a confounder of the association between diabetes incidence and treatment. Simply presenting the crude rates of 2 comparison groups could also lead to biased overall conclusion.

Table 1. Type 2 Diabetes Incidence Rate by Therapy, Age and Gender

	Treated group			Untreated group		
Age-gender group	N = 2294	n (new diabetes case)	% type 2 diabetes	N = 608	n (new diabetes case)	% type 2 diabetes
20-<30 F	30	0	0.0	40	0	0.0
30-<40 F	86	3	3.0	84	4	4.8
40-<50 F	202	18	8.9	74	16	21.6
50-<60 F	236	25	10.6	60	12	20.0
>=60 F	216	70	32.4	16	6	37.5
20-<30 M	36	1	2.8	60	2	3.3
30-<40 M	140	14	10.0	62	8	12.9
40-<50 M	278	36	12.9	92	16	17.4
50-<60 M	488	122	25.0	76	20	26.3
>=60 M	582	301	51.7	44	28	63.6

In table 1, for each age-gender stratum except the subgroup of female aged 20-<30 years, the untreated patients presented higher diabetes incidence than the treated patients, which consistently demonstrated the preventive effort as observed in phase III studies. However, with aging, the incidence increased both for treated and untreated regardless of gender. 66.3% (1522/2294) of the treated patients were 50 years and older versus 32.2% (196/608) of untreated patients, indicating the untreated patients were younger than the treated patients. The younger patients made the overall incidence rate appear lower in the untreated group than the treated group.

SOLUTION: DIRECT STANDARDIZATION METHOD

The purpose of this analysis was to compare the diabetes incidence between the two comparison groups (treated and untreated). However, the exploratory analyses showed that the factors of age and gender were distributed differently between the two comparison groups. As these age and gender factors are also associated with the diabetes incidence, the factors can substantially confound the comparison of the outcome measures between the two comparison groups. Therefore we used the direct standardization method to estimate the age-gender adjusted diabetes incidence in order to adjust for the confounding due to age and gender.

Direct standardization involves 3 major steps. Table 2 begins with 3 columns of raw data and then gives each of the three major steps, First, compute the age-gender specific rate (Column D) for a comparison group (e.g., in Table 2, the untreated group); Secondly, calculate the weighted age-gender specific rates by multiplying a weight (a value between 0-1) (Column E) to obtain Column F; and lastly, the weighted rates are summed leading to the age-gender standardized rate (the bottom number in Column F).

Table 2. Example of Computing Age-gender-standardized Diabetes Incidence Rate for Untreated Patients

Column A	Column B	Column C	Column D	Column E	Column F
Age-gender group	Number of patients	Number of events	Age gender specific rate (%) = column C/column B*100	Weight from a reference population	Weighted rate (%) = Column D * Column E
20-<30 F	40	0	0.0	0.014	0.00
30-<40 F	84	4	4.8	0.041	0.20
40-<50 F	74	16	21.6	0.089	1.92
50-<60 F	60	12	20.0	0.098	1.96
>=60 F	16	6	37.5	0.091	3.41
20-<30 M	60	2	3.3	0.016	0.05
30-<40 M	62	8	12.9	0.057	0.74
40-<50 M	92	16	17.4	0.126	2.19
50-<60 M	76	20	26.3	0.222	5.84
>=60 M	44	28	63.6	0.246	15.65
			Overall crude rate = 18.4	Sum of weights is 1	Sum = 32.0

The weights can come from any one of a number of sources, and in this example was the proportion of the reference patients who fell into a specific age-gender category. The reference patients in this example were the total enrolled patients in the treated group, including patients without sufficient data for clinical evaluation of type 2 diabetes. Applying the same reference population and repeating the same 3 steps above led to the age-gender standardized diabetes incidence rate for the treated group of 25.4%. This example demonstrates the essence of direct standardization is giving weight to each age-gender specific rates and thus the standardized rates are compared between comparison groups as if they have the same age/gender distribution.

In general, the weights can be any set of positive numbers, but here the weights were chosen so that they sum to 1. The general formula for the standardized rate, for weights w_i and age-gender rates r_i, is

$$Adjusted\ Rate = \frac{\sum w_i r_i}{\sum w_i}$$

A normal approximation method was used to test the percentage difference between 2 study groups.

Formula for test statistic:

$$\frac{\text{Difference in \% standardized incidence between treated and untreated}}{\text{Square root of (overall variance of treated + overall variance of untreated)}}$$

The test statistic is compared to quantiles of the standard normal distribution. The age-gender-specific variance for a % of a relatively more frequent outcome (e.g., diabetes) can be estimated from a binomial distribution assumption.

$$Binomial\ \ variance = \frac{(age\text{–}gender\ specific\ \ rate) * (1\text{–}\ age\text{–}gender\ specific\ rate)}{N}$$

, where N is the number of the patients in a particular age-gender group

The age-gender-specific variance for a % (or proportion) of a relatively less frequent outcome (e.g., cancer) can be estimated from Poisson distribution assumption.

$$Poisson\ \ variance = \frac{age\text{–}gender\ specific\ \ rate}{N}$$

, where N is the number of the patients in a particular age-gender group

The overall variance for a group is calculated by the weighted average of the age-gender specific variances, with the weight being the square of the weight used in Column E above:

$$\text{overall variance} = \frac{\sum w_i^2 \times \text{var}_i}{\left(\sum w_i\right)^2}$$

, where Σw_i is 1 in the above example

In this example for untreated patients, a crude incidence rate of 18.4% increased to an age-gender standardized rate of 32.0% mainly because the untreated patients generally had higher incidence rates for older age groups, but fewer patients in those older age groups (table 1), and in the direct standardization process, we gave more weight to those older patients (table 2). This demonstrates that for a disease that occurs more frequently in the older population, the age-gender standardized rate will be higher than the crude one if we apply a reference population that is older than the actual population. Vice versa, the age-gender standardized rate will be lower than the crude one if we apply reference population that is younger than the actual population.

Meanwhile for the treated group, the age-gender standardized rate (25.4%) was almost the same as the crude one (25.7%), with a difference of 0.3%. This small difference occurred because the treated patients taken in the analysis were only those who reported diabetes status (yes/no) in the follow up (N=2294), and they are fewer than the treated reference patients (N=3056) who were enrolled and followed up, thus the age-gender distribution between selected patients in analysis and the reference population was a little bit different (data not shown). But the impact of this age-gender distribution difference is little on the change of rate from the crude one to the standardized one.

As a result, the treated group demonstrated the overall preventive effect of diabetes development as compared with the untreated group in this observational study (25.4 % for treated versus 32.0% for untreated, P=0.014). This overall conclusion is consistent with the approved indication from phase III studies.

DISCUSSION

In this example we illustrated that the direct standardization method can provide a straightforward group comparison and it is easy to understand. Although this example is from a Phase IV study, the method can also be applied to any exploratory or subgroup analyses in phase III studies as long as the age and/or gender function as confounders for the association between therapy and an outcome. In addition, the method is widely used in the epidemiology studies in public health and health claim database analyses for the purpose of age and/or gender adjustment. Note that examination of the rates within each age-gender subgroup is still of value (table 1), while the single number provided by direct standardization is useful as a summary (table 2). In fact, in the hypothetical data in table 1, a potentially stronger drug effect is noted in middle-aged female patients relative to males, which could be examined by other statistical methods (e.g., logistic regression) but it is hidden in the single standardized result for each treatment group.

It is noted that direct standardization is not the only way to account for confounders in clinical studies. It can be considered as one choice along with other statistical methods. The use of direct standardization provides an adjusted rate estimating what would occur with a different age-gender distribution.

Direct standardization method usually refers to the adjustment of confounding effects from the factors age and gender, but the weighting methodology of direct standardization can be also used to adjust for any other potential confounding factors. For example, if a clinical subtype of a disease is associated with both outcome and therapy, thus the confounding effect can be adjusted by using direct standardization for an overall comparison.

DIRECT STANDARDIZATION SAS MACRO

The macro was set up to generate the relevant tables by asking the values for macro parameters for the purpose of generating the tables. Below are the parameter description, example input datasets, example output tables, and an example of macro call that are self-explanatory.

PARAMETER DESCRIPTION

Name	Type	Default	Values Description and Valid Values
INDAT	required		input dataset with the following variables: 1) outcome variable(s) (character, e.g. diabetes (Yes/No)) 2) comparison group variable (character, e.g. therapy (treated, untreated) 3) adjustment variable AGEGRP (character) or AGEGRP and GENDER (character, F or M) or other confounding variables at individual patient record level note: variable attributes for GENDER, AGEGRP or other confounding variable must be the same as those from reference population
			* note: 3 output tables below are optional
OUTDAT1	optional	tb1	output dataset for table1 - crude and standardized rate, standard error, and 95% CIs (note: the dataset name starts with tb1 and continue with outcome variable)
OUTDAT2	optional	tb2	output dataset for table2 - rate difference and P-values from pair-wise comparison for standardized rates (note: the dataset name starts with tb2 and continue with outcome variable)
OUTDAT3	optional	tb3	output dataset for table3 - subgroup rate (i.e., age-specific rate or age-gender specific rate, or specific rate for other confounders) (note: the dataset name starts with tb3 and continue with outcome variable)
			* note: 4 parameters below are for the outcome measure
OUTCOME_VAR	required		outcome variable name from the dataset INDAT
OUTCOME_LB	required		outcome variable label shown on output table
MEASURE_VAR	required		one level of the outcome variable from the INDAT that is chosen for presenting. e.g. the outcome variable DIABETES has 2 levels (Yes and No), 'Yes' is the level chosen for presentation in the output table
MEASURE_LB	required		Label for the level chosen in the output table. - This parameter is set up because the level in the given outcome variable may not be exactly the same as we want to be shown in the output table
			* note: 2 parameters below are for the comparison group
GROUP_VAR	required		the name of comparison variable from dataset INDAT
GROUP_LBLIST	required		Labels for each group to be presented in the output table - since original variable values will be shown in ascending alphabetic order, label must be in the same order as original value - must be enclosed in %str() and separated by ',' (e.g. %str(Treated, Untreated))
			* note: parameters below are for the reference population structure
REFPOP	required		Reference population structure dataset variable RATIO and other variables to be adjusted 1. variable RATIO refers to % of subjects falling into a subgroup sum of ratios should be 1 2. variable AGEGRP, 3. variable GENDER if applicable 4. Other variables to be adjusted
			* note: 2 parameters below are for the adjustment variable(s)
ADJ_VARLIST	required		adjustment variable name from reference dataset REFPOP (e.g, AGEGRP, GENDER and/or other variables)

ADJ_LBLIST	required		adjustment variable name to be shown on output table - must be enclosed in %str() and separated by ',' (e.g. %str(Age Group) or %str(Age Group, GENDER, and/or other confounding variable))
			* note: The macro was created in a way that the comparison of adjusted rates can be made between 2 study groups, each of which has its own age (or age-gender, and/or other confounding variables) specific rates, or the comparison can be made between the study group and a reported point estimate from the reference population. Parameters below are for comparison of rate between study population and reported point estimate of reference population.
RPTRATE	optional		reported prevalence rate from general population survey
RPTVAR	optional		reported variance from general population survey
			* note: parameters below are for choosing data distribution each distribution needs a different formula for computing variance
DIST	optional	B	options: B or b for binomial distributed data P or p for poisson distributed data
			* note: parameters set up for other features of output tables
SUP	optional		suppress CI column on table 1, or p-value comparison of table 2 (CI or ci, P or p)
UNIT	optional	100	unit used to present rate options: 100 (default) 1000 100000
UNITSIGN	optional	%	unit sign to present rate options: %(default) thousandth 100 thousandth
ROUND	optional	.1	rounding for rate (both crude and adjusted) 0.1, 0.01, 0.001, etc
TITLE	required		titles to be given by the user - can start from title3 or title1. But if start from title3, title1 gives mode and title2 gives default output table title by macro
FOOT	required		footnotes to be given by the user

EXAMPLES OF INPUT DATASET

Input dataset for comparison groups

Diabetes	Therapy	Agegrp	Gender
Yes	Treated	<30	F
No	Untreated	30-<60	M
	Untreated	30-<60	M
No	Treated	30-<60	F
Yes	Treated	30-<60	M
Yes	Untreated	>=60	F
...	...	...	...

Reference population dataset (sum of ratios = 1)

Gender	Agegrp	Ratio
f	<30	0.1
m	<30	0.15
f	30-<60	0.3
m	30-<60	0.35
f	>=60	0.06
m	>=60	0.04

SAMPLE OUTPUT TABLES

TABLE 1 -- crude and standardized rate, standard error, and 95% CIs

Outcome	Measure	Group	Crude Rate % (n/N)	Crude Rate 95% CI	Adjusted Rate (%)	Adjusted Rate 95% CI	Adjusted SE
Diabetes	Yes	Therapy1	xxx (xx/xx)	xxx, xxx	xxx	xxx, xxx	xxx
		Therapy2	xxx (xx/xx)	xxx, xxx	xxx	xxx, xxx	xxx
		Therapy3	xxx (xx/xx)	xxx, xxx	xxx	xxx, xxx	xxx

Note: 95% CIs are optional

TABLE 2 -- rate difference and P-values from pair-wise comparison for standardized rates

Therapy 1 and 2	Therapy 1 and 3	Therapy 2 and 3
Diff = xxx	Diff = xxx	Diff = xxx
P = xxx	P = xxx	P = xxx

Note: This table is optional; comparison between group and general population is available and optional

TABLE 3 -- subgroup rate

Outcome	Measure	Group	Overall Crude Rate % (n/N)	Age Group	Gender	Subgroup Rate % (n/N)
Diabetes	Yes	Therapy1	xxx (xx/xx)	<30	F	xxx (xx/xx)
				<30	M	xxx (xx/xx)
				30-<60	F	xxx (xx/xx)
				30-<60	M	xxx (xx/xx)
				>=60	F	xxx (xx/xx)
				>=60	M	xxx (xx/xx)

MACRO CALL EXAMPLES

```
OPTIONS nodate PAGENO=1 orientation=landscape ls=132 ps=46 missing='';
ods rtf file = "directory\filename";

   %stand  (indat =  exampdat,

            outdat1 = ,
            outdat2 = ,
            outdat3 = ,

            outcome_var = Diabetes,
            outcome_lb = Diabetes,
            measure_var = Yes,
            measure_lb = Yes,

            group_var = therapy,
            group_lblist = %str(treated, untreated),

            adj_varlist = agegrp gender,
            adj_lblist = %str(Age Group, Gender),

            refpop = stand,

            rptrate = ,
            rptvar = ,
            dist = B,
            sup = CI,
            unit = ,
            unitsign = ,
            mode = RMT,
            round = .01,
            title = %str(title3...),
            foot = %str(footnote1... ));

   %stand  (... for another outcome);
   ...
   so on

ods rtf close;
```

CONCLUSION

Direct standardization method can present rates adjusted for confounders (e.g., age and gender) and provide a valid and straightforward comparison between groups. The method can be applied to data mining where comparison groups are not balanced in confounders. The macro developed here can facilitate the use of the direct standardization method.

REFERENCES

John Last. A dictionary of epidemiology (Third edition). Page 35 Confounder definition. Oxford. 1995

Robert N. Anderson, Harry Rosenberg. Age standardization of death rates: implementation of the year 2000 standard. National vital statistics reports 47(3) Oct 7, 1998

Breslow NE and Day NE . Statistical methods in cancer research: volumn II: the design and analysis of cohort studies (IARC scientific publication, No82). Chapter 2. April 1994

Boyle P and Parkin DM. Cancer registrations: principles and methods. Chapter 5. Statistics methods for registries. Page 126-158. IARC scientific publication No. 95

ACKNOWLEDGMENTS

The authors are grateful to Yanhong Li, from inVentiv Clinical Solutions LLC. for implementing the programming and validating.

CONTACT INFORMATION

For requesting a copy of the direct standardization macro, contact the author at:

Daojun Mo
Eli Lilly and Company
Indianapolis, IN, 46225
Work Phone: 317-433-3236
E-mail: mo_daojun@lilly.com