%ME: A SAS® Macro to Assess Measurement Equivalence for Patient-Reported Outcome Measures

Qinlei Huang, St. Jude Children's Research Hospital, Memphis, TN

ABSTRACT

Patient-reported outcomes (PROs) are the consequences of diseases, their treatment, or both as reported by patients. PRO measures are important for evaluating new treatments, drugs, biological agents, and devices when used as effectiveness end points in clinical trials. Measurement equivalence is a function of the comparability of the psychometric properties of the data obtained via mixed modes (e.g., paper-and-pencil questionnaires, web-based questionnaires); in paired data (e.g., patient-reported outcome, parent-reported outcome); and/or at different time points. Measurement equivalence must be established prior to using a PRO measure for further statistical analysis and modeling. Multiple statistical methods have been developed to test measurement equivalence by evaluating means (TOST, TTest), variances (Levene's test), or correlations and agreements (Pearson Product Moment Correlation, Intra-class correlation, Weighted Kappa, and Spearman's Rho). A SAS Macro %ME has been written to provide all the above mentioned statistics for measurement equivalence automatically and concurrently. A sample with real data will be provided for illustration. Keywords: Measurement equivalence, survey methodology

INTRODUCTION

There are multiple statistical tests to assess measurement equivalence by evaluating means, variances, agreements, and correlations. The statistical methods assessing measurement equivalence fall into two groups based on different study design: parallel groups data and clustered data.

For the parallel groups study design, patients are randomly assigned to one of two study arms. Patients in each study arm will complete one survey. For a comparison study of PRO vs. ePRO, patients in one study arm would complete the original paper version of the PRO measure, and patients in the other arm would complete the ePRO measure. This paper focuses on data using parallel groups study design.

Another way to administer a survey via mixed modes is to have the same subjects complete multiple surveys using different modes. For the clustered data study design, each patient will answer both the paper and the online surveys. The most common strategies used to examine these kind of clustered data include paired *t*-tests, Pearson product-moment correlation coefficients, Spearman's Rho test, intra-class correlation, and weighted kappa statistics. Another manuscript will introduce these methods and the corresponding SAS macro `%agreement` in detail.

Mode	Subject
Mode 1 (e.g., paper-and-pencil survey)	1 … n_1
Mode 2 (e.g., web survey)	1 … n_2
…	
Mode j	1 … n_j

Table 1a. survey administered via mixed modes: parallel group data

Subject	Mode 1	Mode 2	…	Mode j
1	1	1	1	1
…	…	…	…	…
n	n	n	n	n

Table 2b. survey administered via mixed modes: clustered data

STATISTICAL METHODS TO TEST MEASUREMENT EQUIVALENCE

There are two types of statistical methods used to test the equality of measurements. One is based on means; another is based on variances.

MEASUREMENT EQUIVALENCE BASED ON MEANS

The following statistical methods are frequently used to test the equality of measurements based on means: independent t-test, Cohen's d effect size estimate, and TOST equivalence test.

Independent *t*-test

The independent t-test provides inference statistics to test the difference between means. The hypotheses for an independent t-test are as follows:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

A rejection of the null hypothesis suggests that the means of the two groups are significantly different. Please note that acceptance of the null hypothesis does not directly suggest the means of the two groups are the same. Independent t-tests are implemented by using SAS PROC TTEST.

```
proc ttest data=a;
 var Measure;
 class class;
run;
```

Below are SAS Output generated by a PROC TTEST statement, including: statistics, confidence limits, statistics for t-test, and statistics for equality-of-variance test.

```
class           N         Mean      Std Dev      Std Err     Minimum      Maximum
1              63      28.5556     10.6525       1.3421      10.0000      50.0000
2              63      28.4603     10.8417       1.3659      10.0000      49.000
Diff (1-2)              0.0952     10.7475       1.9149


class         Method             Mean        95% CL Mean        Std Dev      95% CL Std Dev
1                               28.5556     25.8727  31.2384     10.6525      9.0631  12.9233
2                               28.4603     25.7299  31.1908     10.8417      9.2240  13.1528
Diff (1-2)    Pooled            0.0952      -3.6950   3.8854     10.7475      9.5602  12.2743
Diff (1-2)    Satterthwaite     0.0952      -3.6950   3.8854

Method            Variances       DF    t Value    Pr > |t|
Pooled            Equal          124      0.05       0.9604
Satterthwaite     Unequal     123.96      0.05       0.9604

Equality of Variances
Method        Num DF    Den DF    F Value     Pr > F
Folded F         62        62       1.04       0.8902
```

Output 1. Output from a PROC TTEST Statement

The equality of variances test is an F-test having the null hypothesis that two normal populations have the same variance. If the *p*-value is smaller than α (which is typically 0.05), then the same variance hypothesis is rejected and, thus, the t-test statistics should be calculated by using the Satterthwaite method because of unequal variances. If the *p*-value is larger than α (which is typically 0.05), then the same variance hypothesis is supported and, thus, the t-test statistics should be calculated by using the pooled method because of equal variances.

If the *p*-value of a t-test is smaller than α (which is typically 0.05), then the difference between the means is significant.

Cohen's d effect size estimate

Cohen's *d* is an appropriate estimate of effect size for the comparison between two means. It indicates the standardized difference between two means and can be used, for example, to accompany reporting of t-tests. It is also widely used in meta-analyses.

```
data cohend_1;
  set ttest_1;
  retain std1 . std2 . n1 . n2 . cohend .;
```

2

```
  if class='1' then do;
   std1=stddev; n1=n;
  end;
  if class='2' then do;
   std2=stddev; n2=n;
  end;
  if class='Diff (1-2)' then do;
   cohend=mean/(sqrt(((n1-1)*(std1**2)+(n2-1)*(std2**2))/(n1+n2)));
  end;
  drop std1 std2 n1 n2;
run;
```

TOST equivalence test

Two one-sided tests (TOST) is a statistical method proposed by Schuirmann (1987) that involves conducting two one-sided *t*-tests for the mean or mean difference. The hypotheses for an equivalence test are as follows:

$H_0: \mu < \theta_L \text{ or } \mu > \theta_U$

$H_1: \theta_L \leq \mu \leq \theta_U$

where θ_L and θ_U are the lower and upper bounds specified in the TOST option of the PROC TTEST statement and μ is the analysis criterion (mean, mean ratio, or mean difference, depending on the analysis). The equivalence test is conducted by performing two separate tests:

$H_0: \mu < \theta_L \quad H_1: \mu \geq \theta_L$

and

$H_0: \mu > \theta_U \quad H_1: \mu \leq \theta_U$

TOST equivalence test is implemented by using the TOST option for SAS PROC TTEST procedure:

```
proc ttest data=work.a tost(2.18);
 var Measure;
 class class;
run;
```

The overall *p*-value is the larger of the two *p*-values of those tests. The hypothesis of equivalence (H_1) is supported if the *p*-value is smaller than α (which is typically 0.05). Rejection of H_0 in favor of H_1 at significance level α occurs if and only if the 100(1-2α)% confidence interval for μ is contained completely within (θ_L, θ_U). So, the 100(1-2α)% confidence interval for μ is displayed in addition to the usual 100(1-α)% interval.

MEASUREMENT EQUIVALENCE BASED ON VARIANCES

Levene's test

Levene's test is an inferential statistic used to assess the equality of variances in different samples. It tests the homogeneity of variances of measurements against a null hypothesis that the population variances are equal. Levene's test is considered to be the standard homogeneity-of-variance test. A Levene's test is implemented by using the HOVTEST option for SAS PROC GLM:

```
proc glm data=a;
  class class;
  model Measure=class/solution;
  means class/hovtest=levene;
run;
```

If the resulting *p*-value of Levene's test is less than some critical value (typically 0.05), then the null hypothesis of equal variances is rejected.

MACRO

INSTALLATION

The macro introduced here, %me, is actually a group of smaller macros (`%me`, `%ttest`, `%tost`, `%hovtest`, `%report_ttest`, `%report_tost`, `%report_hovtest`, and `%report_me`). To use it within a SAS program, put these macros (all found in the ME directory) into a central directory (e.g., c:\SAS\me). Then add the following to the SAS program before calling `%me`:

3

```
%let meroot = c:\SAS\ME;
options mautosource sasautos=("&meroot",sasautos);
```

This tells SAS to look at the contents of the &meroot directory to find new macro definitions (in particular, `%me` and its component macros).

PARAMETERS

Here, I explain how to use `%me` and its component macros `%ttest`, `%tost`, `%hovtest`. Examples are provided for each macro. Some of the parameters might best be understood via the examples that follow.

```
%me(datain= ,
    varlist ,
    n= ,
    class= ,
    fix= ,
    bound= ,
    data_ttest= ,
    data_tost= ,
    data_hovtest= );
```

- datain: the name of the input SAS dataset, including the name of the SAS library where the input data set is located.

- varlist: the list of variables to be tested for measurement equivalence, each separated by a space.

- n: the total number of variables to be tested for measurement equivalence.

- class: a dichotomous variable indicating the groups upon which the measurement equivalence test will be based. Note: please code the class variable as a numeric variable with values 1 or 2.

- fix: an index variable indicating whether the bound for the TOST is fixed; if the bound is a fixed number, then fix=1; otherwise fix=0.

- bound: if bound is a fixed number, then bound is equal to the fixed number; if bound is standard deviation times a coefficient, then bound is equal to the coefficient.

- data_ttest: the name of the output SAS dataset for the results of *t*-test and Cohen's d effect size, including the name of the SAS library where the output data set is located.

- data_tost: the name of the output SAS dataset for the results of TOST equivalence test, including the name of the SAS library where the output data set is located.

- data_hovtest: the name of the output SAS dataset for the results of Levene's test, including the name of the SAS library where the output data set is located.

EXAMPLES

Our examples will be based on an artificial dataset having four variables: M1, M2, M3, and class. M1-M3 are measurement variables, and class is a dichotomous variable with '1' indicating web-based survey and '2' indicating paper-and-pencil survey. Table 2 shows the first five observations of the dataset.

Obs	M1	M2	M3	Class
1	28	38	11	2
2	25	21	11	1
3	21	30	11	2
4	22	22	14	1
5	24	15	19	2

Table 2. Dataset for measurement equivalence test (first five observations)

Example 1: `%ttest`: a SAS macro for independent *t*-test

Independent *t*-test provides inference statistics to test the difference between means. A SAS macro `%ttest` has been written to generate a summary report of the results of independent tests on a bunch of interested variables. Cohen's d effect size is also included in the summary table.

```
%ttest(datain=work.a,
       varlist=%str(M1 M2 M3),
       n=3,
       class=%str(class),
       data_ttest=work.ttest);
```

The SAS macro `%ttest` first performs *t*-tests by using PROC TTEST on a SAS dataset defined by `datain`.
`varlist` defines outcome variables; `n` defines the total number of outcome variables; and `class` defines the
class variable in the class statement of PROC TTEST. `%ttest` then uses SAS/ODS and SAS/DATA to read the
results of *t*-tests based on the results of the equality-of-variances tests; to calculate the Cohen's d effect size; and to
store values of the corresponding n, mean, standard deviation, lower and upper 95% confidence interval of mean, t-
value, degree-of-freedom, *p*-value, and Cohen's d into the output dataset defined by `data_ttest`.

The above example runs *t*-tests of SAS datasets in the work library. There are three outcome variables `M1`, `M2`, `M3`
and a dichotomous class variable `class`. The output SAS dataset `ttest` is stored in the work library. A summary
table (Table 3) is produced by using `%report_ttest` in the central directory based on the output dataset `ttest`.

| Variable | Class | N | Mean | Std. | 95% L | 95% U | Independent t-test ||||
							t-value	df	P(t)	Cohen's d
M1	1	49	23.73	2.99	22.88	24.59				
	2	44	22.82	3.80	21.66	23.97				
	Diff (1-2)		0.92	3.40	−0.49	2.32	1.30	91	0.198	0.27
M2	1	49	28.08	10.99	24.93	31.24				
	2	44	27.18	10.84	23.89	30.48				
	Diff (1-2)		0.90	10.92	−3.61	5.40	0.40	91	0.692	0.08
M3	1	49	14.73	6.63	12.83	16.64				
	2	44	13.89	5.51	12.21	15.56				
	Diff (1-2)		0.85	6.13	−1.68	3.38	0.67	91	0.507	0.14

**Table 3. Summary table of results of *t*-tests and Cohen's d effect size estimations generated by SAS macro
`%ttest`**

Example 2: `%tost`: a SAS macro for TOST equivalence test

Two one-sided tests (TOST) is a statistical method involving two one-sided *t* tests for the mean or mean difference.
TOST equivalence test is implemented by using the TOST option for SAS PROC TTEST procedure. A SAS macro
`%tost` has been written to generate a summary report of the results of TOST equivalence tests on a set of variables-
of-interest.

```
%tost(datain=work.a,
      varlist=%str(M1 M2 M3),
      n=3,
      class=%str(class),
      fix=0,
      bound=0.2,
      data_tost=work.tost);
```

The SAS macro `%tost` first produces TOST equivalence tests by using the TOST option in the PROC TTEST
procedure on a SAS dataset defined by `datain`. `varlist` defines outcome variables; `n` defines the total number of
outcome variables; and `class` defines the class variable in the class statement of PROC TTEST. `fix` indicates
whether the bound is a fixed number (1) or a coefficient (0)[1]. `bound` defines the upper and lower bounds for the
TOST tests. Then, `%tost` uses SAS/ODS and SAS/DATA to read the results of TOST tests based on the results of
the equality-of-variance tests and to store the values of the corresponding n, mean, standard deviation, bound, lower
and upper 90% confidence interval of mean, *p*-value, and assessment into the output dataset defined by `data_tost`.

The above example runs TOST equivalence tests on an SAS dataset in the work library. There are three outcome
variables `M1`, `M2`, `M3` and a dichotomous class variable `class`. The `bound` is not a fixed number. Instead, it is
equal to 0.2*standard deviation (i.e., [−0.2*std.dev, 0.2*std.dev]). The output SAS dataset `tost` is stored in the work

---

[1] There are two typical ways to define the bounds used for a TOST equivalence test: 1) using a fixed number; 2) having bound
equal to standard deviation times a coefficient.

library. A summary table (Table 4) based on the output dataset `tost` is produced by using `%report_tost` in the central directory.

Variable	Class	N	Mean	Std.	TOST Equivalence Test				
					Bound	90% L	90% U	P(t)	Assessment
M1	1	49	23.73	2.99					
	2	44	22.82	3.80					
	Diff (1-2)		0.92	3.40	0.68	−0.26	2.09	0.631	Not equivalent
M2	1	49	28.08	10.99					
	2	44	27.18	10.84					
	Diff (1-2)		0.90	10.92	2.18	−2.87	4.67	0.287	Not equivalent
M3	1	49	14.73	6.63					
	2	44	13.89	5.51					
	Diff (1-2)		0.85	6.13	1.23	−1.27	2.96	0.382	Not equivalent

Table 4. Summary table of results of TOST equivalence tests generated by SAS macro `%tost`

Example 3: `%hovtest`: a SAS macro for Levene's test

Levene's test is an inferential statistic used to assess the equality-of-variances in different samples by evaluating the homogeneity-of-variances of measurements and is considered to be the standard homogeneity-of-variance test. Levene's test is implemented by using the HOVTEST option for SAS PROC GLM procedure. A SAS macro `%hovtest` has been written to generate a summary report of the results of Levene's tests on a set of variables-of-interest.

```
%hovtest(datain=work.a,
        varlist=%str(M1 M2 M3),
        n=3,
        class=%str(class),
        data_hovtest=work.hovtest);
```

The SAS macro `%hovtest` first produces Levene's tests by using the HOVTEST option in the PROC GLM procedure on a SAS dataset defined by `datain`. `varlist` defines outcome variables; `n` defines the total number of outcome variables; and `class` defines the class variable in the class statement of PROC GLM. Then, `%hovtest` uses SAS/ODS and SAS/DATA to read the results of Levene's tests and to store the corresponding n, mean, standard deviation, minimum, 25% quartile, median, 75% quartile, maximum, and *p*-value for the Levene's test into the output dataset defined by `data_hovtest`.

The above example runs Levene's tests on a SAS dataset in the work library. There are three outcome variables `M1`, `M2`, `M3` and a dichotomous class variable `class`. The output SAS dataset `hovtest` is stored in the work library. A summary table (Table 5) based on the output dataset `hovtest` is produced by using `%report_hovtest` in the central directory.

Variable	Class	N	Mean	Std.	Min	25%	Median	75%	Max	*p*-value (Levene's)
M1		93	23.30	3.41	12.00	22.00	23.00	26.00	28.00	0.13
	1	49	23.73	2.99	15.00	22.00	24.00	26.00	28.00	
	2	44	22.82	3.80	12.00	20.50	22.50	26.00	28.00	
M2		93	27.66	10.87	10.00	18.00	28.00	37.00	50.00	0.89
	1	49	28.08	10.99	10.00	20.00	26.00	37.00	50.00	
	2	44	27.18	10.84	10.00	16.50	28.00	36.50	49.00	
M3		93	14.33	6.11	10.00	10.00	12.00	15.00	44.00	0.54
	1	49	14.73	6.63	10.00	10.00	12.00	17.00	44.00	
	2	44	13.89	5.51	10.00	11.00	12.00	14.00	34.00	

Table 5. Summary table of results of Levene's tests generated by SAS macro `%hovtest`

Example 4: `%me`: a SAS macro for measurement equivalence test

A SAS macro `%me` has been written to generate a summary report of the results of all equivalence tests mentioned above: independent *t*-test, Cohen's d effect size estimate, TOST equivalence test, and Levene's test.

```
%me(datain=work.a,
    varlist=%str(M1 M2 M3),
    n=3,
```

```
class=%str(class),
fix=0,
bound=0.2,
data_ttest=work.test,
data_tost=work.tost,
data_hovtest=work.hovtest;
```

The SAS macro `%me` first calls the component macros `%ttest`, `%tost`, and `%hovtest` and produces output SAS datasets `data_ttest`, `data_tost`, and `data_hovtest`. Then, `%me` uses SAS/DATA to produce a comprehensive SAS dataset `data_me` based on the three datasets mentioned above. Dataset `data_me` stores the statistics of exploratory analysis, including n; mean; standard deviation; the statistics of *t*-test, consisting of lower and upper 95% confidence interval of mean, t-value, degree-of-freedom, and *p*-value; Cohen's d effect size estimate; the statistics of TOST equivalence test, including bound, lower and upper 90% confidence interval of mean, *p*-value, and assessment; and *p*-value for Levene's test. Lastly, `%me` calls `%report_me` in the central directory to produce a summary table (Table 6) based on the output dataset `data_me`.

| Variable | Class | N | Mean | Std. | Independent *t*-test | | | | | Effect Size | TOST Equivalence Test | | | | | Levene's Test |
					95% L	95% U	t-value	df	P(t)	d	Bound	90% L	90% U	P(t)	Assessment	P(F)
M1	1	49	23.73	2.99	22.88	24.59										
	2	44	22.82	3.80	21.66	23.97										
	Diff (1-2)		0.92	3.40	−0.49	2.32	1.30	91	0.198	0.27	0.68	−0.26	2.09	0.631	Not equivalent	0.13
M2	1	49	28.08	10.99	24.93	31.24										
	2	44	27.18	10.84	23.89	30.48										
	Diff (1-2)		0.90	10.92	−3.61	5.40	0.40	91	0.692	0.08	2.18	−2.87	4.67	0.287	Not equivalent	0.89
M3	1	49	14.73	6.63	12.83	16.64										
	2	44	13.89	5.51	12.21	15.56										
	Diff (1-2)		0.85	6.13	−1.68	3.38	0.67	91	0.507	0.14	1.23	−1.27	2.96	0.382	Not equivalent	0.54

Table 6. Summary table of results of measurement equivalence tests generated by SAS macro `%me`

CONCLUSION

SAS macro %me, is actually a group of smaller macros (`%me`, `%ttest`, `%tost`, `%hovtest`, `%report_ttest`, `%report_tost`, `%report_hovtest`, and `%report_me`). It produces measurement equivalence tests, including the independent *t*-test, Cohen's d effect size estimate, TOST equivalence test, and Levene's test of variance equality, in a reliable and automatic way. By using SAS/Macro, SAS/SQL, SAS/ODS, and SAS/DATA, %me provides a sleek summary report of all statistical analyses mentioned above.

REFERENCES

Chang P-C and Yeh C-H. 2005. Agreement between Child Self-Report and Parent Proxy-Report to Evaluate Quality of Life in Children with Cancer. *Psycho-Oncology* 14(2): 125-134.

Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, Lenderking WR, Cella D, and Basch E. 2009. Recommendations on Evidence Needed to Support Measurement Equivalence between Electronic and Paper-Based Patient-Reported Outcome (PRO) Measures: ISPOR ePRO Good Research Practices Rask Force Report. *Value in Health* 12(4): 419-429.

Cronk BC and West JL. 2002. Personality Research on The Internet: A Comparison of Web-based and Traditional Instruments in Take-home and In-class Settings. *Behavior Research Methods, Instruments, & Computers* 34(2): 177-180.

De Beuckelaer A and Lievens F. 2009. Measurement Equivalence of Paper-and-Pencil and Internet Organisational Surveys: A Large Scale Examination in 16 Countries. *Applied Psychology* 58(2): 336-361.

Schuirmann DJ. 1987. A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* 15(6): 657-680.

ACKNOWLEDGMENTS

biostatisticians Catherine Billups, Yinmei Zhou, and Wei Liu for their valuable comments on my work.

RECOMMENDED READING

- SAS/STAT(R) 9.2 User's Guide, Second Edition
- SAS® Certification Prep Guide: Advanced Programming for SAS® 9
- SAS® For Dummies®

CONTACT INFORMATION

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

Name: Qinlei Huang
Enterprise: St. Jude Children's Research Hospital
Address: 262 Danny Thomas Place, Mail Stop 723
City, State ZIP: Memphis, TN 38103
Work Phone: 901.595.2027
Fax: 901.595.4585
E-mail: qinlei.huang@stjude.org
Web: www.stjude.org

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.