Precision Study for a Qualitative Assay

Paul Hshieh*, Ph.D.
Paul.hshieh@fda.hhs.gov

Tie-Hua Ng*, Ph. D.
Tiehua.ng@fda.hhs.gov

US Food and Drug Administration, 1401 Rockville Pike, Rockville, MD 20852

Abstract

Percent agreement of a test result to that of a reference method is often used in reporting the analysis of a precision study for a qualitative assay. However, the interpretation of such an analysis is unclear when less than perfect agreement is observed. In this paper, we propose a pairwise agreement approach and discuss how it works.

Key Words: Pairwise agreement, pairwise comparisons, precision study, repeatability, reproducibility, intermediate precision, random pairing, independent pair.

Introduction

Although the statistical methods for analyzing the precision study for a quantitative assay have been well established (see NCCLS), it is not clear how a precision study for a qualitative assay ought to be analyzed.

One approach commonly used for a qualitative assay is based on the percent agreement of the test results as compared with a reference method. A 100% agreement is the best possible outcome; however, the interpretation of the results becomes difficult when less than perfect agreement is observed because percent agreement with the known result does not provide information about sources of variation and their releative importance.

In this paper, pairwise agreement is proposed to analyze the precision study for a qualitative assay. We first present a study design of a precision study and then discuss how our proposed pairwise agreement approach works. Since the most common qualitative tests we see are serological, qualitative outcomes will be referred to as reactive and non-reactive but the results can apply to any binary outcome.

Study Design (Precision)

Let us consider a study design of a precision study for a qualitative assays using one kit lot. A precision study is conducted at 3 laboratory sites with three operators at each site. Each operator performs 2 runs per day over 5 days with two replicates per run. It should

* The views expressed in this paper are not necessarily of the U.S. Food and Drug Administration.

be noted that (i) operators are nested in sites, (ii) days are nested in operators and (iii) runs are nested in days. The study design is shown in Table 1.

Table 1. Study Design

<table>
<tr><td colspan="9">Site (one lot)</td></tr>
<tr><td colspan="3">1</td><td colspan="3">2</td><td colspan="3">3</td></tr>
<tr><td colspan="9">Operators</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr>
<tr><td>X</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="5">Site 1, Operator 1</td><td colspan="4">Site 1, Operator 1, Day a</td></tr>
<tr><td colspan="5">Day</td><td colspan="4">Run</td></tr>
<tr><td>a</td><td>b</td><td>c</td><td>d</td><td>e</td><td colspan="2">1</td><td colspan="2">2</td></tr>
<tr><td>X</td><td></td><td></td><td></td><td></td><td>A</td><td>B</td><td>C</td><td>D</td></tr>
</table>

Percent Agreement Approach: Total Pairwise Comparisons

To assess the precision of the assay, we will estimate the percent agreement within-run, between runs, days, operators and sites. Based on the conditions of repeatability and reproducibility defined in **CLSI-EP5-A2** (see the Appendix), we considered (i) the within-run agreement as repeatability, (ii) site-to-site agreement as reproducibility and (iii) run-to-run, day-to-day and operator-to-operator agreements as intermediate precisions.

The following definitions will be used in this paper.

- Percent agreement = (number of pairs with agreement) / (total number of pairwise comparisons)
- Run-pair: pair of runs for run-to-run comparison
- Day-pair: pair of days for day-to-day comparison
- Operator-pair: pair of operators for operator-to-operator comparison
- Site-pair: pair of sites for site-to-site comparison
- $c(n,k) = n!/[(n-k)!k!]$

To compute the percentage of agreement, we need to count the number of pairwise comparisons and the number of agreement for within-run, between runs, days, operators and sites. The numbers of pairwise comparisons for repeatability, intermediate precisions and reproducibility are determined as follows.

1. Within Run (repeatability):
   There are 90 pairs within run pairwise comparisons (3 operators x 3 sites x 5 days x 2 runs).

2. Between Runs (intermediate precision):
   There are 45 run-pairs (3 operators x 3 sites x 5 days). For each run-pair, there are 4 pairwise comparisons (AC, AD, BC and BD). Therefore, there are a total of 180 (= 45 x 4) pairwise comparisons for between runs intermediate precision.

3. Between Days (intermediate precision):
   There are 90 day-pairs (3 operators x 3 sites x c(5,2)). For each day-pair, there are 16 pairwise comparisons (4 run-pairs x 4 pairs/run-pair). Therefore, there are a total of 1,440 (= 90 x 16) pairwise comparisons for between days intermediate precision.

4. Between Operators (intermediate precision):
   There are 9 operator-pairs (3 sites x c(3,2)). For each operator-pair, there are 400 pairwise comparisons (25 day-pairs x 4 run-pairs/day-pair x 4 pairs/run-pair). Therefore, there are a total of 3,600 (= 9 x 400) pairwise comparisons for between operators intermediate precision.

5. Between Sites (reproducibility):
   There are 3 site-pairs. For each site-pair, there are 3,600 pairwise comparisons (9 operator-pairs x 25 day-pairs/operator-pair x 4 run-pairs/day-pair x 4 pairs/run-pair). Therefore, there are a total of 108,000 (= 3 x 3,600) pairwise comparisons for reproducibility.

Percent Agreement Approach: Independent Pairwise Comparisons

It is not clear how to compute the confidence interval for the percent agreement based on total pairwise comparisons, except within run comparison, since all possible pairs are not independent. To resolve this issue, we propose random sampling to form independent pairs. Such a random pairing is discussed for between runs and days intermediate precisions in the followings:

1. Between Runs (intermediate Precision)

   For each of the 45 run-pairs, there are 4 possible pairwise comparisons but only 2 independent pairwise comparisons can be formed. We randomly form 2 independent pairwise comparisons for each of the 45 run-pairs resulting in 90 independent pairwise comparisons. **For example,** there are 4 possible pairwise comparisons (AC, AD, BC and BD; see Table 1). Thus, we have two ways to form 2 independent pairs, which are (i) AC, BD and (ii) AD, BC.

   However, the percent agreement depends on random pairing. To overcome this problem, we average the results over (i) and (ii), keeping the number of independent pairwise comparisons unchanged. **For example,** assuming A and C are reactive and B and D are non-reactive then the result is 2/2 for (i) and 0/2 for (ii). The average is ½.

2. Between Days (Intermediate Precision)

   Step1: For each of the 9 operators, there are 10 (=c(5,2)) day-pairs, but only 2 independent day-pairs can be formed. We randomly form 2 independent day-pairs, resulting in 18 independent day-pairs. It should be noted that there are 15 [= c(5,2) x c(3,2)/2] ways to form 2 independent day-pairs. For each of 18 independent day-pairs, we randomly form 2 independent run-pairs resulting in 36 independent run-pairs. In addition, there are two ways to form 2 independent run-pairs. For example, for day-pair ab (see the table below), the two independent run-pairs are (i) aR1bR1, aR2bR2 and (ii) aR1bR2, aR2bR1.

Day			
a		b	
Run		Run	
I	II	I	II
aR1	aR2	bR1	bR2

   Step 2: For each of the 36 independent run-pairs, randomly form 2 independent pairs for comparisons resulting in 72 independent pairwise comparisons. It should be noted that there are two ways to form 2 independent pairs for comparisons. Again, the percent agreement depends on random pairing. This issue can be resolved by average the results over (1) 15 ways of forming 2 independent day-pairs, (2) two ways of forming 2 independent run-pairs, and (3) two ways of forming 2 independent pairs for comparisons, keeping the number of independent day-pairs, run-pairs, and pairwise comparisons, respectively, unchanged.

Discussion and Further Research

A similar approach can be applied for Between Operators and Between Sites. If the number of replicates is 3 or more, a similar approach may be used for repeatability; however, if there are more than 2 runs per day and/or 2 replicates per run, then more steps will be needed.

The confidence interval can be constructed based on Binomial distribution using random pairing; this is problematic as the result may change if another random pairing is made. Averaging over different ways of forming independent pairings resolves this issue. However, it is unlikely that a confidence interval based on binomial distribution is valid. Furthermore, the numerator might not be an integer.

In addition, counting the number of agreement in the calculation of the percent agreement is complicated and tedious. Further research is needed.

Appendix. Precision Studies (CLSI-EP5-A2; NCCLS)

Repeatability Conditions: conditions where independent test results are obtained with the same method on identical test material in the same laboratory by the same operator using the same equipment within a short interval of time

Reproducibility Conditions: conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment

Reference

NCCLS. Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline-Second Edition. NCCLS document EP5-A2 [ISBN 1-56238-542-9]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2004