

SIR Comparison Limited Application Protocol for Determining Reliability When
a SIR Program is Updated (Reliability Comparison Protocol)
May 31, 2016

Disclaimer: This Protocol may only be used for SIR methods with a Third Party Evaluation and listed on the NWGLDE website. It is not suitable when modifications are made to SIR methods; those must be re-evaluated by a third party. To determine if this protocol is applicable for the intended use, the user must consult the NWGLDE prior to using this SIR protocol and receive guidance from the NWGLDE concerning its applicability.

Purpose:

The purpose of this protocol is to determine the **reliability** of two copies of the same method; i.e., to ensure that the two copies produce the **same results** when they are run on different types of computer, are written in different computer languages, or run under different operating systems. Thus it is important to establish that this is a “reliability” protocol and not a “validity” protocol.

The two terms, reliability and validity, are often used interchangeably when they are not related to statistics. When critical readers of statistics use these terms, however, they refer to different properties of the statistical or experimental method.

Reliability is another term for consistency. If one person takes the same personality test several times and always receives the same results, the test is said to be *reliable*.

A test is **valid** if it measures what it is supposed to measure. If the results of a personality test showed that a very shy person was in fact very outgoing, the test would not be *valid*.

Reliability and validity are independent of each other. A measurement may be valid, but not be reliable, or be reliable but not valid. If a bathroom scale was reset to read 10 pounds lighter than a person’s true weight, then the scale would **reliable** if the weights were the same each time a person steps on it, but not **valid** because the weights would always be wrong.

An independent Third Party Evaluation according to an NWGLDE acceptable protocol is used to certify the **validity** of a statistical method; i.e., it certifies that a particular method can find a product release (a leak) with a probability of a false alarm of 0.05 or less and a probability of detection of .095 or greater. This “Reliability Protocol”, **does not certify** the **validity** of a SIR method, and thus can only be used on methods which have already been evaluated with an approved SIR protocol and listed on the NWGLDE List.

Methodology:

This Reliability Comparison must be performed by a party independent of the SIR vendor with no financial interest in the SIR vendor’s company whose method is being examined. The party performing the comparison must be someone familiar with statistical analytical methods and have access to a computer (or computers) capable of running the copies of the SIR program being evaluated. (Henceforth

in this document this party will be referred to as the 'evaluator'.) The evaluator may be asked to provide information to the NWGLDE which supports his suitability to perform the Reliability Comparison.

Steps:

1. The evaluator must have at least 200 monthly inventory data (SIR datasets) from tanks. Datasets do not have to be from 200 different tanks, but do have to be discrete data periods. One hundred of these datasets will be used in the comparison.

Datasets should be from tanks that had passing SIR results (determined to be tight) as well as from tanks that had failing or inconclusive SIR results (determined to have leaks or were inconclusive due to data variability or data error). Approximately 50% of the datasets used in the comparison would be from tanks with No leak (Pass); and approximately 25% of the datasets in the comparison used must be from Product Loss or Product Gain (Fail) SIR results and approximately 25% of the datasets from Inconclusive or SIR results.

2. The evaluator will obtain an executable copy of the original SIR program from the vendor with instructions on how to load and operate the SIR program on the evaluator's computer (or computers). This copy must be the originally listed NWGLDE SIR version with no modifications from the full evaluation performed when the method was evaluated. The evaluator will also obtain an executable copy of the original SIR program written in the different language.

The vendor must certify to the evaluator that the copy being compared is the same program and not merely the same program with a different name, or an entirely different SIR program.

3. If the newest copy of the SIR program is to be run on a specific operating system (i.e. Windows or Mac) the vendor must specify that to the evaluator. If the system was developed on one platform such as Windows, but the vendor desires it to be able to operate on other platforms such as Apple, OS X, or Linux, this protocol should be used to demonstrate that the SIR program is capable of producing identical results on multiple operating systems.
4. The evaluator will analyze the selected data on each copy of the program and tabulate the statistics produced by each copy (see Table, below). (If for any reason an analysis fails to run to completion on any dataset, the dataset may be discarded and another substituted until there are 100 total results. The discarded dataset may not be used in Step 5 below; use the substituted dataset in Step 5.)

The evaluator must use the same 100 monthly SIR datasets successfully used on both copies.

5. The evaluator will prepare a table comparing the results of Step 4. Table 1 with an example (*in italics*) may be used as a model.
6. The evaluator will compare the results of the two copies. The copies may be considered to be comparable if the results conform to the following:

Quantitative Comparisons: All quantitative comparisons (TH, MDL, and CLR) must agree to at least two places beyond the decimal. A discrepancy greater than 0.01 between any of the three quantities will result in the evaluator having to declare the two copies to be different.

Qualitative Comparisons: The two copies must be in 100% agreement **on the qualitative variable (Pass, Loss, Gain or Inconclusive)**. Any disagreement will result in the evaluator having to declare the two versions to be different.

Table 1 (Example)
Variables to be Compared

Original Version								Copy					
Test	SITE	TANK #	TH	MDL	CLR	P/F/I		SITE	TANK #	TH	MDL	CLR	P/F/I
1	1	001	0.086	0.173	0.065	P		1	001	0.086	0.173	0.065	P
2	1	002	0.019	0.105	0.029	P		1	002	0.019	0.105	0.029	P
3													
4													
5													
6													

Key to Abbreviations

TH = Threshold

MDL = Minimum Detectable Leak Rate

CLR = Calculated Leak Rate

P/F/I =monthly SIR result as determined by SIR analysis

- The evaluator may prepare a report confirming the results of his comparison. The report must contain at a minimum, a summary of the work done, conclusions reached, and the table of the results obtained from his comparison as shown in the table above. The comparison must be signed and dated by the evaluator and should be on the evaluator’s letterhead. The evaluator may include additional information at the request of the SIR vendor or the NWGLDE.

As an alternative to Steps 2. and 3. above, should an evaluator lack the appropriate computer infrastructure to do the reliability protocol at his place of business or his home, it may be acceptable for the evaluation to be done at a different location. If a different location is needed, the evaluator should discuss this with the NWGLDE prior to attempting the reliability protocol. The evaluator’s report (item 7. above) must state where the reliability protocol was performed and whether multiple computers and/or what equipment was used as well as describe any other deviation allowed by the NWGLDE.