Cleaning Memo for February 2019

Significant Figures: Back to Basics

Those of you that have taken any of my training courses or webinars know that I have certain "pet peeves". One of them is *significant figures*, particularly in expressing sampling recovery factors. For example, it is common that I see in company reports and in published papers where recovery percentages are given to three or four decimal places. For example, a swab recovery study is reported as 73.468%. Well, swabbing is like a type of manual cleaning, and the numbers are just not that precise. There are several elements that contribute to this misuse. One is that often the input data is just two or three significant figures. So if the measured amount per swab in the recovery sample is 127 µg and the amount spiked is 150 µg, the recovery is reported as 84.667%. And then if there are three replicates, for example with recoveries of 84.667%, 78.667%, and 87.333%, the average is reported as 85.556%. Another reason good scientists report numbers that way is probably because they use an Excel spreadsheet for calculations, and the spreadsheet is set up to report to three (or four) decimal places.

According to the well accepted rules involving significant figures, if the input data is only three significant figures, the reported result of the calculation should be rounded to three significant figures. Now there is an exception to this in that if the result of a calculation is used in subsequent calculations, the unrounded result may be used in those subsequent calculations, but the *final* result should be reported to the appropriate number of significant figures. Furthermore, it is generally accepted that while the unrounded number may be used for subsequent calculations, what is *reported* should be reported with the proper number of significant figures. I believe the reason for this is that if a number is used for a large number of subsequent calculations (for example, ten subsequent calculations) and if rounding is done at *each* calculation, the final result conceivably could be "off" by a digit or more in the final result (at the end of the tenth calculation.)

A second, and what I consider a more significant reason, for not reporting a sampling recovery percentage to three or four decimal places, is that regardless of the number of significant figures in the input data used to calculate the percentage, the resulting percentages, as a *practical* matter, are just *not that precise*. In the example given previously with percentage values from three replicates, the best that I could have any confidence in would be two significant figures (that is, with no digits to the right of the decimal). So I would report the average (my "official" recovery percentage) as 86%. If the "official" recovery percentage was to the lowest of the three replicates, then that would be 79%.

Some might object that rounding up (that is rounding 85.556% up to 86%) makes it easier to pass in an executed protocol, and that we should really be using the lower value of 85.556%, or even use "rounding down" principles to use 85%. My belief is that if the difference between 86% and 85% recovery is the difference between passing and failing in a protocol, then I should really design my cleaning process to be more robust.

Copyright © 2019 by Cleaning Validation Technologies. This copyright protected Cleaning Memo may be printed for research, compliance and scientific purposes. Any other use, including downloading of the file and including commercial distribution, is illegal and unethical. (February 2019 Cleaning Memo)

Some might further argue that this discussion is like asking "how many angels can fit on the end of a pen?" If I have a robust cleaning process, it doesn't matter how many significant figures I report. While there is some truth to that, it does reflect a "casualness" to well established scientific practices.

There are some good sources for us in the pharmaceutical industry to utilize in addressing significant figures and rounding in the use and reporting of data. One is the Section 4 of Volume III of the FDA's "ORA Laboratory Manual" on statistics and data presentation (https://www.fda.gov/downloads/ScienceResearch/FieldScience/LaboratoryManual/UCM092179.pdf). Here is what that manual says about significant figures:

Definitions and Rules for Significant Figures

- All non-zero digits are significant.
- The most significant digit in a reported result is the left-most non-zero digit: 359.741 (3 is the most significant digit).
- If there is a decimal point, the least significant digit in a reported result is the rightmost digit (whether zero or not): 359.741 (1 is the least significant digit). If there is no decimal point present, the right-most non-zero digit is the least significant digit.
- The number of digits between and including the most and least significant digit is the number of significant digits in the result: 359.741 (there are six significant digits).

And here is what the FDA manual says about significant figures in *calculated* results:

Most analytical results in ORA laboratories are obtained by arithmetic combinations of numbers: addition, subtraction, multiplication, and division. The proper number of digits used to express the result can be easily obtained in all cases by remembering the principle stated above: numerical results are reported with a precision near that of the least precise numerical measurement used to generate the number. Some guidelines and examples follow.

Addition and Subtraction

The general guideline when adding and subtracting numbers is that the answer should have decimal places equal to that of the component with the least number of decimal places:

21.1
2.037
6.13
29.267 = 29.3, since component 21.1 has the *least* number of decimal places

Multiplication and Division

The general guideline is that the answer has the same number of significant figures as the number with the fewest significant figures:

$$\frac{56 \times 0.003462 \times 43.72}{1.684}$$

A calculator yields an answer of 4.975740998 = 5.0, since one of the measurements has only two significant figures.

And here is what the FDA manual says about rounding:

The following rules should be used:

- If the extra digit is less than 5, drop the digit.
- If the extra digit is greater than 5, drop it and increase the previous digit by one.
- If the extra digit is five, then increase the previous digit by one if it is odd; otherwise do not change the previous digit.

Copyright © 2019 by Cleaning Validation Technologies. This copyright protected Cleaning Memo may be printed for research, compliance and scientific purposes. Any other use, including downloading of the file and including commercial distribution, is illegal and unethical. (February 2019 Cleaning Memo)

Another good source is the USP General Notices on “Rounding Rules”. It differs slightly from the FDA approach for rounding in that the third and fourth bullet points would be replaced with something like “If the extra digit is *5 or greater than 5*, drop it and *increase* the previous digit by one”.

What is discussed here probably has no significant effect on the outcome of a protocol. However, if we become sloppy using one well-established scientific practice, it is likely in the future that this sloppiness could carry over to other well-established practices.

Copyright © 2019 by Cleaning Validation Technologies. This copyright protected Cleaning Memo may be printed for research, compliance and scientific purposes. Any other use, including downloading of the file and including commercial distribution, is illegal and unethical. (February 2019 Cleaning Memo)