5-2020

Equivalency Testing for Two Formulations of a Clinical Laboratory Control Material

Jessica M. Hart
University of Nebraska Medical Center

Equivalency Testing for Two Formulations of a Clinical Laboratory Control Material

Jessica Hart

Abstract:

Clinical laboratory control materials are an integral part of legally-mandated and highly regulated quality control protocols in all clinical laboratories. These controls ensure accurate performance of the laboratory testing and instrumentation used to produce medical test results for millions of patients. It is of clinical and public health interest to ensure the diagnostic test results which affect so many people are regulated by the most accurate and precise controls.

Formulation changes in control materials have the potential to impact laboratory quality control. In this study, data from two formulations of a hematology control were compared to assess equivalency of the mean results for each parameter. The Two One-Sided T-Test of equivalency was used to compare the old formulation of this hematology control to a new formulation. Equivalency testing assumes a null hypothesis that two sets of data differ, which is contrary to other statistical methods which assume a null hypothesis of non-difference.

Data for 15 hematology parameters were gathered for 6 different lots of product, on 4 different instruments, with 3 lots each for the old and new formulations. ANOVA analysis revealed significant differences in means between lots and instruments, so equivalency testing was stratified on lot and instrument to obtain specific details about any lack of equivalency. Testing was also done with non-stratified data from the mean results of the old and new formulations to obtain general conclusions about the equivalency of each parameter in each level.

The results of the TOST analysis described in this report show that all except two of the parameters in the new formulation are equivalent to the old formulation. The parameters which did not demonstrate equivalence between the two formulations were white blood cell (WBC) count and red blood cell size distribution (RDW). Although the lack of equivalency in these two parameters is notable, equivalency testing of the means alone is not sufficient to fully determine the clinical significance. Both accuracy and precision of controls must be maintained after a formulation switch. TOST equivalency testing is able to assess accuracy by determining whether target concentrations in the new formulation are equivalent with target concentrations in the old formulations but does not address any potential changes in precision as a result of altered variability. While overall equivalency of the mean results is an important step in evaluating the accuracy of the new formulation, the manufacturer should also consider evaluating the equivalence of variance to verify precision before making any decision to switch to the new formulation.

Introduction:

The Clinical Laboratory and its Role in Public Health:

Before a medical condition can be treated, it must be diagnosed. Over the course of human history, medicine has made significant advancements in the ability to identify and treat diseases. In the modern era, and with the help of the scientific method, our diagnostic capabilities have become more and more sophisticated. Advances in chemistry, biology, automation, human physiology, and medicine have presented vast potential for diagnostic testing methods. The field of clinical laboratory science first emerged in the early 1900s (Kotlarz, 1998). In the present day,

clinical laboratory science is an invaluable component of the healthcare industry, providing laboratory test results to healthcare providers who use the information to make diagnostic and treatment decisions for millions of patients. Clinical laboratory results inform many medical decisions, such as wellness checks, preventative medicine, early detection and diagnosis, and personalized treatment plans based on individual genetic makeup (American Clinical Laboratory Association, 2014). Accuracy of these laboratory results is critical to the success of healthcare teams. Quality control protocols, proficiency testing, and maintenance of employee and lab accreditation, certification, and licensure are all necessary to maintain optimal laboratory performance.

Clinical laboratories and public health laboratories work in conjunction to obtain diagnostic data from populations. While clinical laboratories focus on individual diagnosis, public health laboratories focus on epidemiologic and population diagnosis, watching for patterns in illnesses and diseases (Wilson et al., 2010). Laboratory test methods are often the same, and quality control measures must be in place to ensure accuracy of test results. In the first few months of 2020, the global public health system's reliance on laboratory testing was highlighted, front and center, by the sudden surge in demand for appropriate diagnostic testing for the SARS-CoV-2 coronavirus (Sheridan, 2020). Test methodology is constantly evolving, and new diseases will continue to emerge into epidemics and pandemics which can only be quantified by effective laboratory testing protocols.

The same regulatory bodies that govern public health laboratories are also responsible for governing clinical laboratories. The U.S. Food and Drug Administration (FDA) is responsible for clearing medical diagnostic devices and test kits for approved use and categorizing clinical laboratory tests according to complexity (U.S. Food and Drug Administration, 2020). The

Centers for Disease Control and Prevention (CDC) provides research, technical assistance, and advice to clinical laboratories to improve quality standards (Centers for Disease Control and Prevention, 2018). These institutions also work with the Department of Health and Human Services (DHHS) and local public health departments to maintain a high level of laboratory testing and reporting quality. All of these institutions recommend a quality control protocol, which is used to verify the accuracy of diagnostic results by testing a "mock" patient sample with pre-determined results. This is also known as a clinical laboratory control sample.

The Importance of Clinical Laboratory Controls:

Clinical laboratory quality control materials are samples used to validate whether laboratory instrumentation and testing procedures are functioning according to pre-defined specifications. These samples are manufactured to mimic patient samples in both physical appearance, and qualitative and quantitative results. In the field of clinical laboratory diagnostics, quality control is a legally mandated and highly regulated system for validating the accuracy of tests performed on patient samples. In 1967, and again in 1988, the Public Health Services Act was amended with the Clinical Laboratory Improvement Amendments (CLIA), requiring clinical laboratories to have a quality control program in place (Clinical Laboratory Improvement Amendments of 1988). The World Health Organization (WHO) has also published a set of Good Clinical Laboratory Practices (GCLP) which urge clinical laboratories to maintain quality control procedures and to subscribe to additional accreditation agency guidelines, such as the College of American Pathologists (CAP) and the American Society of Clinical Pathology (ASCP) (Ezzelle et al., 2007; World Health Organization, 2009). Under these guidelines, daily (or day-of-use) quality control is required for every individual test system. For quantitative tests, control materials must be run at two or three levels within the reportable range of the test, or at levels

where clinical decisions are made. Levels of the control are meant to cover the entire reportable range of a test system. In other words, there should be a "low" level control to check the test system's capability to accurately produce low results, and there should be a "high" level control to check the accuracy of high results. There are often "normal" levels in between as well, to test the accuracy of results which are in the normal, or average, range of patient results (Bio-Rad Laboratories, 2018).

To monitor test system performance, the results of each control test are compared to an acceptable "assayed" range, given by the manufacturer of the control (Njoroge & Nichols, 2014; Medical Laboratory Observer, 2015). Because clinical laboratories must often rely on other manufacturers to provide them with quality control samples, it's highly important for control material to vary as little as possible from lot to lot, even when the formulation is changed by the manufacturer. When clinical laboratories run a test on a quality control sample and obtain a result within the pre-specified acceptable range, they can be confident their instrument is functioning properly. (Njoroge & Nichols, 2014). Laboratories rely on these control materials to check the accuracy of their instruments, keeping logs of every control test and using Levey-Jennings charts, Excel spreadsheets, instrument logs, or inter-laboratory quality control databases to monitor data for shifts or trends which might indicate changes in instrument or test system performance (Bio-Rad Laboratories, 2008). For this reason, control materials must remain consistent over time, providing customers with a reliable baseline to gauge their test system performance. It must be assured that the ability of customers to make clinical decisions, or perform quality control, will not be altered by any changes in material formulation.

Once a test platform is validated using a clinical laboratory control sample, it will be used to test true patient samples and provide clinical decision-makers with accurate results (Medical Laboratory Observer, 2015). The impact of diagnostic test results on patient healthcare is

significant. An estimated 35% to 70% of all patient treatment encounters are informed by laboratory test results (Ngo et al., 2016; American Clinical Laboratory Association, 2014; Centers for Disease Control and Prevention, 2018). These laboratory results are then used in a myriad of clinical diagnoses and treatment decisions, meaning their accuracy is of vital importance to the healthcare quality of millions of patients. Not only could one laboratory mistake mean the difference between life or death for an individual patient, but a single malfunctioning instrument can produce inaccurate results for hundreds, or even thousands, of patients. Thankfully, the strict regulation and monitoring of quality control in clinical laboratories reduces the number of diagnostic errors from laboratory results. Error rates in the analytical phase of testing account for 13-32% of errors surrounding laboratory testing in the United States (Bonini et al., 2002). The rest of these errors occur either before or after the actual laboratory test is performed. These include labeling errors, patient mix-ups, incorrectly ordered tests, etc. Consistent performance of clinical laboratory control materials is a vital component of the healthcare industry.

Equivalency Testing:

In the clinical setting, equivalency testing is commonly used for testing pharmaceuticals in clinical trials. To assess the bioequivalence or function of two different drugs (or two different formulations of a similar drug), equivalency testing provides a statistical method for ruling out any significant differences between the two drug formulas. As opposed to difference testing (such as a t-test), where the null hypothesis is that there is *no* statistically significant difference between samples, the null hypothesis for equivalency testing is that there *is* a statistically significant difference between two samples (Limentani et al., 2005). When the study aim is to rule out significant differences between two samples, equivalency testing is more appropriate than traditional difference testing, and can be used in a variety of scenarios such as in assessing

two manufactured products which are meant to have the same function (Dixon et al., 2018). Traditional difference testing is only able to identify the presence of a significant difference between two groups. It is unable to identify the presence of significant equivalence between two groups. Equivalency testing, on the other hand, can identify significant equivalence between two groups, since the null hypothesis assumes a lack of equivalence.

Some examples of the practical usefulness of difference testing are clinical trials in which one treatment method is being compared to another. The question of interest in this these studies is whether switching to a new treatment would result in significant improvement in patient outcomes, versus the old treatment. An example of practical usefulness for equivalency testing is drug testing, where two drugs are tested for the same use. In this case, the question of interest is whether both drugs are "bioequivalent," that is, to determine whether one drug has the same treatment effect as the other drug.

Early equivalency testing was devised by determining an interval of maximum acceptable difference between two samples, and comparing the actual measured difference to the acceptable range. In this method, (usually referred to as Westlake's confidence interval method) the samples were considered to be equivalent if the confidence interval of the true difference in means fell entirely within the acceptable range. However, this method used assumed non-difference in the null hypothesis, rather than non-equivalence. Later methods began to switch to assuming non-equivalence with the interval method, using a student's two-sided t-test to determine if the average difference in means of two samples fell outside equivalence limits (Ialongo, 2016).

Today, the most common form of equivalency testing is the Two One-Sided T-Test, or TOST. In this method, a lower bound and upper bound are combined to create an equivalence range, where results falling inside the range are considered to be equivalent (Lakens, 2017). A

parametric one-sided t-test is performed for both the upper and lower bounds, and two separate p-values are calculated and compared to the test's alpha. Because two separate t-tests are done as part of this test (one for the inferiority test and one for the superiority test), each t-test has the same alpha - usually 0.05. Equivalency is determined when both the hypothesis of inferiority (the difference in means is smaller than the lower bound) and the hypothesis of superiority (the difference in means is larger than the upper bound) are both disproved simultaneously (Ialongo, 2016).

An acceptable equivalence criterion (upper and lower bounds) must be defined as part of the TOST study design. This equivalence range is determined from prior knowledge, previous data, or relevance to the particular application of the data. (Dixon et al., 2018). If the difference in mean values falls within this equivalence range, the null hypothesis is rejected, and the two samples are concluded to be statistically equivalent. (Ialongo, 2016). In the case of clinical laboratory controls, the equivalence range is determined by the clinical significance of any difference between samples. Clinical significance (also called clinical relevance) can be difficult to define, because it is determined by answering questions about whether patients are being helped or harmed in a meaningful way. Clinical significance is often determined by how a difference in results can change the way healthcare providers utilize or interpret those results in regard to treatment options (Armijo-Olivio, 2018). Differences which are statistically significant may not always have clinical relevance, and differences which are clinically significant may not always be statistically significant.

In the case of clinical laboratory hematology controls, significance is mostly determined by whether a difference in results will cause a change in the acceptability of those results, per the laboratory's quality control protocol (Medical Laboratory Observer, 2015). If differences in

control materials cause changes to the interpretation of quality control, this can have an impact on the testing process (and therefore patient results) as well.

The manufacturer of the hematology control in this study has determined acceptable limits for variability in the mean results of each parameter. These acceptable limits of variability in results for the laboratory end-users have been used as the equivalence ranges for the TOST equivalency testing in this study. They are manufacturer-specific information, determined from previous control data over many years, prior to the manufacture of the new formulation. Upon changing the formulation of this control, the results must remain within the acceptable limits to be determined statistically equivalent with the old formulation.

Study Design, Data, and Statistical Analysis:

Materials and Study Design:

This data was collected from a local manufacturer of hematology controls for clinical laboratory diagnostic use. To protect proprietary information, specifics about production and formulation are not discussed in this paper, but this product is a laboratory control which resembles human blood and produces hematology results which fall into expected target ranges on automated hematology instruments.

Because hematology testing involves the measurement of multiple parameters of bloodwork, such as white blood cell count and differentials, red blood cell count, hemoglobin levels, platelet counts, etc., this hematology control is manufactured with 15 quantifiable parameters. All 15 parameters in the new hematology control formulation were compared to all 15 parameters from the old hematology control formulation through equivalency testing.

The main objective of this study was to use TOST statistical equivalency testing to compare the mean results of the new formulation with the mean results of the old (current) formulation. Through manufacturing, three pilot lots of this hematology control were created with a new formulation. These three pilot lots were compared to three lots of the original formulation of the same type of hematology control.

Data:

Data were collected using automated hematology instrumentation (also called hematology analyzers), which is the clinical laboratory industry standard. The hematology control of interest is manufactured specifically for use on automated hematology instruments. To encompass instrument-to-instrument variability, four different automated hematology instruments were used to collect data for this study.

This hematology control contains 3 levels (low, normal, and high) to evaluate the performance of hematology instrumentation when measuring low, normal, and high results. All three levels were compared separately between the old and new control formulations. For each of the three lots from the old formulation, and each of the three lots from the new formulation, data was collected over a period of 104 days (the period of time between when the product is shipped to customers and when the product expires). To eliminate variability due to differences in product age, this period of time was dependent on the date of product manufacture. This means testing was staggered in such a way that data points were collected at the same time points for each lot throughout aging.

It is not uncommon for automated hematology instruments to have occasional errors or misreads while performing a test. These most commonly occur due to bubbles in the sample or

tubing, or insufficient volume for testing (short-sampling), and can easily be identified by individuals who are qualified to operate these instruments and who are familiar with how these errors can affect quantitative results (Keohane et al., 2020). Errors such as these have been removed from the dataset by a trained individual and treated as missing values in the data analysis.

Categorical variables:

- Group: The three lots created with the old formulation were combined into the "Old" group, while the three pilot lots created with the new formulation were combined into the "New" group.
- Lot: Each individual lot has been identified according to the Julian date when it was manufactured.
- Level: This control is divided into three levels: (1=Low, 2=Normal, 3=High)
- Instrument: Each of the four automated hematology instruments (or hematology analyzers) have been assigned a letter from "A" to "D" to protect industry-sensitive information.

Continuous variables:

This hematology control offers quantitative measurement for the following 15 parameters:

- WBC: White blood cell count
- WBC Differential: 5 types of white blood cells are differentiated (out of 100%) -
    - Neut %: Neutrophil percentage
    - Lymph %: Lymphocyte percentage

    - Mono %: Monocyte percentage
    - Eos %: Eosiniophil percentage
    - Baso %: Basophil percentage
- RBC: Red blood cell count
- HGB: Hemoglobin
- HCT: Hematocrit
- MCV: Mean red blood cell volume
- MCH: Mean red blood cell hemoglobin
- MCHC: Mean red blood cell hemoglobin concentration
- RDW: Red blood cell size distribution
- PLT: Platelet count
- MPV: Mean platelet volume

Statistical Methods:

Data was exported from the manufacturer's database and organized into a separate file for data analysis using Microsoft Excel and SAS software version 9.4.

Basic descriptive analysis was done to count the frequency of samples for each categorical variable. Table 1 illustrates the number of samples within each level of each lot, run on the four different instruments. The total number of samples per lot, per group, and per level is tallied in Table 1, as well as the overall total number of samples in the study. Means and standard deviations have also been calculated for all 15 hematology parameters in Table 2. Graphical and visual examination of the data, as well as the Shapiro-Wilk test of normality, were used to

determine whether the data in each level of each parameter met the normal distribution assumption.

The data was separated according to Levels 1, 2, and 3. To determine whether stratification by lot or instrument would be necessary, the lot-to-lot and instrument-to-instrument differences were evaluated using ANOVA, then a 2-way ANOVA model was used to assess whether the average results of each lot differed for each instrument (lot and instrument interaction). If these lot-to-lot and instrument-to-instrument differences were found to be significant at a significance level of $p \leq 0.05$, comparisons between the old and new formulations were made for individual lots, stratified by instrument. In other words, on each instrument, each lot from the new formulation was compared to an equivalence range calculated from the old formulation. For each parameter, an equivalence range was determined using the mean of the old group and a (+/-) range deemed clinically appropriate from the manufacturer (see Table 19 for equivalence ranges). Using these ranges, one-sample equivalency testing was performed on each parameter of each new lot, compared to the average of all three old lots, using the Two One-Sided T-Test of equivalence (TOST).

In addition to stratifying the comparisons based on lot and instrument, one-sample equivalence tests were done for each parameter in the new formulation based on the average result over all lots and instruments. The average for each level of each parameter was compared to the target equivalence ranges listed in Table 19. Finally, two-sample equivalency testing was also performed to directly compare the averages of the new and old formulations without stratification by lot and instrument.

For each comparison, the sample from the new formulation was said to be equivalent to the old formulation if both t-tests within the TOST procedure simultaneously resulted in p-values

≤0.05. For the equivalence tests done over all strata of lots and instruments, if comparisons from a lot were equivalent for all four instruments, the lot was designated “Equivalent” to the old formulation at that level. If at least one of the comparisons from one of the instruments was not equivalent, the lot was designated “Not Equivalent” to the old formulation at that level. For the one- and two-sample testing without stratification, a parameter was said to be “Equivalent” to the old formulation if all three levels were significantly equivalent. If at least one level of the parameter was not significantly equivalent, the parameter was said to be “Not Equivalent” to the old formulation. A correction method was not used for this study. It was deemed more desirable to test equivalence of parameters individually to understand specific details regarding lot and instrument information about those which lacked evidence for statistical equivalency.

Results:

Upon graphical and visual examination, stratified samples appeared to meet the normality assumption. After testing the normality assumption for each level of each parameter using the Shapiro-Wilk test of normality, some groups appeared not to meet the normality assumption. Visual and graphical inspection of the data showed differences between lots and instruments within each level. By stratifying each level by lot and instrument, these smaller groups were tested for normality. Several of these groups met the normality assumption, although several others visually appeared to have a normal distribution but did not meet the statistical level of significance for the Shapiro-Wilk test. Equivalency testing was deemed appropriate despite slight deviations from normality in data distribution, due to the large sample size. When sample sizes are large, violation of normality is often still acceptable (Ghasemi & Zahediasl, 2012).

ANOVA results revealed significant differences between lots and instruments in the new formulation, so analysis was stratified based on lot and instrument. After stratification, most of the parameters were found to be statistically equivalent between the new and old formulations using TOST for equivalence. Monocytes, eosinophils, basophils, red blood cell counts, hemoglobin, hematocrit, mean red blood cell volume, mean red blood cell hemoglobin, mean red blood cell hemoglobin concentration, and platelet counts were equivalent over all instruments, lots, and levels, with significant p-values of <0.001 in nearly all cases. For level 3, all 15 hematology parameters were equivalent for the new and old formulations, on all lots and all instruments.

Five of the parameters exhibited inequivalence to some degree in the stratified one-sample testing. In the new formulation, white blood cell counts in level 2 were higher than the target equivalence range, resulting in insufficient evidence for equivalence. White blood cell counts in levels 1 and 3 were equivalent. Neutrophil percentages were nearly equivalent, but the sample mean from lot 9007, run on instrument B, was too high to fall into the equivalency range.

Lymphocyte percentage was not equivalent for both levels 1 and 2 in the stratified one-sample testing. Again, the sample means from lot 9007 on instrument B fell slightly out of the target equivalence range – this time lower than expected. Red blood cell size distribution was consistently inequivalent between the old and new formulation in level 1, and also inequivalent for lot 8274 in level 2. This inequality was present regardless of which instrument was used to test the samples. Lastly, the mean platelet volume was inequivalent for a few instances in level 1. Although the means didn't fall outside of expected equivalence ranges, the p-values were insignificant and there was insufficient evidence to determine equivalency.

All results of the one-sample equivalency testing, stratified by lot and instrument, are shown in Tables 3 – 17. Each table contains target values and target equivalency ranges, as well as calculated means and p-values, stratified by level, lot, and instrument. Interpretations of the p-value for each stratum are included in the tables. White blood cell parameters, such as the white blood cell count, neutrophil percentage, and lymphocyte percentage, seemed to be prone to inequivalence for lot 9007 when run on instrument B. For red blood cell size distribution and mean platelet volume, there appeared to be no noticeable pattern due to lots or instruments.

After one- and two-sample testing without stratification, two parameters lacked equivalency. After averaging results over all lots and instruments, the neutrophil percentage, lymphocyte percentage, and mean platelet volume were found to be equivalent to the old formulation in all levels. The white blood cell count and red blood cell size distribution were not equivalent to the old formulation. White blood cell count was higher than the acceptable range for level 2, and red blood cell distribution was lower than the acceptable range for level 1. Table 18 contains results for one-sample equivalency testing without stratification, and table 19 contains results for two-sample equivalency testing without stratification.

Discussion:

Explanation of the results

A correction method for multiple comparisons was not deemed necessary in the case of this study. Rather than making an overall conclusion about the equivalence of the formulations, it was preferable to identify the individual components which did not show equivalence between the old and new formulations. This provides specific information about adjustments to each

parameter that could be made in product manufacturing to achieve equivalence with the old formulation if desired - whether increasing or decreasing target concentrations or fine-tuning the formulation itself. All except two of the parameters in the new formulation were equivalent with the old formulation. White blood cell count and red blood cell distribution were inequivalent in at least one level. Three of the parameters which were deemed "Not Equivalent," in the stratified testing were equivalent in all strata except for one or two. For example, neutrophil and lymphocyte percentage only failed to be equivalent for lot 9007 when run on instrument B (but not for all 3 levels), meaning the lack of equivalence in those cases was likely due to a combination of instrument and lot variation, rather than the product formulations or target concentrations themselves. In these instances, neutrophil percentage was higher than the target average (calculated from the average of the old formulation), while lymphocyte percentage was lower than the target average. However, in the non-stratified testing, these three parameters were equivalent for all three levels.

Red blood cell size distribution was clearly lower in level 1 of the new formulation than in the old formulation; in level 2, it was higher only in lot 8274. This parameter is a measure of the variability in red blood cell volume for the whole population of red blood cells. The most likely reason for the narrower red blood cell size distribution in level 1 was the source of red blood cell material. The three lots of the new formulation were made with different ratios of a new type of red blood cells combined with an old type of red blood cells. It is slightly unexpected that the new type of red blood cells would have more uniform size than the old type of red blood cells in level 1, but more variable size in level 2, especially when combining the new and old types together. This may simply be a result of the biological variability in the

batches of red blood cells used for each lot, because many individual packs of red blood cells must be combined to create the necessary volumes.

White blood cell count may be the parameter with the most notable lack of equality between formulations. The white blood cell count only lacked equivalency in level 2, because the target counts in the three new formulation lots were too high to be equivalent to the average counts in the three old formulation lots. Because this pattern isn't seen in levels 1 or 3, this does not appear to be an issue with the new white blood cell formulation itself. The most likely reason for this lack of equality is a difference in target concentration set by the manufacturer when preparing the new formulation lots.

The manufacturer has established acceptable ranges for target concentrations of each count-based parameter. When creating a lot of product, concentrations are deemed acceptable if they fall within these ranges, which are also called the product specification ranges. The specification ranges may be wide enough to allow for differences in concentration between lots to be inequivalent via the TOST equivalency testing method. Unfortunately, the wide range in acceptable parameter concentrations makes equivalency testing with TOST insufficient to decide whether the inequality of average results between the two formulations is clinically significant. By testing for equivalency between the mean results of each parameter in the old formulation versus new formulation, this analysis only assesses whether target concentrations in the new formulation lots are equivalent with target concentrations in the old formulation lots.

Variation in parameter values naturally occurs from differences in the manufacturing process, such as changes in manufacturing equipment and chemicals, differences between processing techs, or from random biological variability in the raw materials. These sources of variability are nearly impossible to control. To minimize this variation, protocols are written with

very strict guidelines, and the acceptable manufacturing targets should fall within the specification ranges that have been pre-determined for each level of the control. Variation also naturally occurs for the end-users in the laboratories which run the controls, due to differences in techs, equipment, and automated instrumentation. An imprecisely manufactured control would add variability to a quality control process which is already subject to variability. For this reason, additional analysis for equivalence of variance is desirable to supplement the findings in this study.

Minimizing variability is crucial for laboratory quality controls, but it is difficult to eliminate all variables in the manufacturing and testing of biological materials. Four different instruments were used for testing and three different lots were produced for both the old and new formulations to account for some of this variability. Multiple laboratory technologists ran these samples on different days during the 104-day testing period, which represented the dating between shipment and expiration when the product would be in the hands of customers. Different lots were meant to capture some of the variability that occurs during manufacturing, and different instruments were used to capture some of the variability that occurs during end-user testing. From a statistical standpoint, it would be ideal for any new formulation to be tested in many more lots, on many more instruments, and in many different laboratory locations, but resources are limited. It is very expensive and time- and labor-intensive to produce and test pilot lots, so the common practice of this manufacturer is to produce three lots for data collection.

The Testing Method

Equivalency testing was originally chosen as the statistical analysis method for this study because the original question for these two formulations of control material was “Can the new formulation be a functionally and clinically equivalent replacement for the old formulation?” In

other words, would this new formulation appear to be the same product as the old product, once in the hands of clinical laboratory scientists who are using it to perform quality control? Would these laboratory scientists even be able to notice a difference?

TOST equivalency testing of means alone may not be enough to answer these questions fully. Both accuracy and precision are important traits of control material results. TOST can only detect significant equivalency between means, so it can answer whether a new formulation of a control maintains the same level of accuracy as an old formulation. Variability testing would be necessary to determine whether a new formulation maintains the same level of precision as an old formulation. In some ways, precision may be a more important measure of comparison between two formulations, because target concentrations can easily be adjusted by the manufacturer as long as they fall within specification ranges. Clinical laboratory scientists monitor their quality control results every day, watching for shifts, trends, or outliers in the data. They recognize that lot-to-lot variability exists and may compensate for some of the changes that occur when switching lots, but a complete change in control product would be a large disruption to their carefully monitored system. When introducing a new formulation, customers may have some flexibility in adjusting target means, but less flexibility in adjusting for large changes in precision.

Because of this, in addition to testing the means of each parameter in the old and new formulations, the variance should also be compared between the old and new formulations. This would most likely be done with an equal-variance test, such as an F-test, Levene's test, or Bartlett's test. In these tests, the variances of the new and old formulations would be compared, using the null hypothesis that they are equal versus the alternative hypothesis that they are not. Again, this rests the burden of proof on the difference of the variances, rather than the equality

(Li, 2015). Variance is a measure of precision, rather than accuracy. To evaluate differences in variance between the old and new formulations would be to ensure results from the new formulation were just as precise as the old formulation. This is valuable to clinical laboratories, who base their acceptance criteria for a single quality control result on whether or not it remains within a certain distance from the mean (usually either within assay ranges provided by the manufacturer or within 2 standard deviations of the mean established by the lab itself).

Equivalency testing isn't the only method which would have produced useful results. Difference testing may have also sufficed, but it has weaknesses when answering the questions above. Very large sets of data occasionally have the issue of being statistically overpowered (Ialongo, 2017). This means there is an increased risk of type I error, or the risk of falsely concluding a statistically significant difference exists when it truly does not. Additionally, difference testing would have only allowed for a lack of evidence of a difference, rather than evidence of equivalence. Because equivalence is desired between the two formulations of this control, difference testing would not have provided enough certainty to make the appropriate conclusion.

However, equivalency testing is not without its disadvantages. A necessary part of equivalency testing is determining the appropriate equivalency ranges, which can prove challenging if the meaning of "equivalence" is to be maintained as something significant. Clinical significance, or the ability of a formulation change to have a real impact on the eventual treatment of patients in a healthcare setting, is the main concern in this case. The equivalence ranges have been determined through manufacturing expertise and a close relationship with clinical laboratories in the field, to determine the maximum acceptable level of difference before clinical laboratories begin to question the quality of the results.

Conclusions and Future Steps

The new formulation was equivalent to the old formulation for the majority of parameters. All except two of the parameters in the new formulation of this hematology control had mean results which were equivalent to the mean results of the old formulation.

White blood cell counts (WBC) for all three new lots were manufactured with a target value for level 2 that was different than the target value for the old lots, while still remaining acceptable for the manufacturing specifications. However, for the equivalency testing, the difference in target concentrations resulted in insufficient evidence for equivalence.

For the second parameter, which was red blood cell size distribution (RDW), it is impossible to control the level of red blood cell size variability from incoming raw materials. Using material from a single vendor or single shipment might improve uniformity, but further studies would be required to see if this is true.

Moving forward, the TOST method of equivalency testing can be used to verify the statistical equivalence of results before and after formulation changes, if consistency in mean results is desired by the manufacturer. However, future studies of this nature should be accompanied by equivalence of variance testing as well, to capture the equivalence of both accuracy and precision between old and new formulations of controls. Target values and assay ranges are communicated to customers with each lot. Significant changes in the target values for any parameter, even if in product specifications, might raise concern from end-users. As long as the product's precision is maintained and customers continue to recover target values within assay ranges, quality control in laboratories can continue to meet requirements and regulations.

As a vital component of the healthcare industry, laboratory control manufacturers have an obligation to provide customers with the most consistent products possible. In turn, clinical laboratories have an obligation to provide accurate and precise results to healthcare providers, so

decisions regarding patient treatment can be well-informed. Adverse health events occur when conditions are improperly diagnosed due to laboratory error, which is why the government and regulatory organizations have mandated a quality control system for the clinical laboratory. Without reliable laboratory controls, the quality of our clinical laboratory system would deteriorate, which is why determining the equivalence of these control formulations is so important to the healthcare industry and the lives of millions.

Appendix A: Tables

Table 1: Frequencies by Group, Lot, Instrument, and Level

Group	Lot	Instrument	Level 1 (N)	Level 2 (N)	Level 3 (N)	N total
Old	7324	A	30	30	30	90
		B	30	31	30	91
		C	31	30	30	91
		D	30	30	30	90
			121	**121**	**120**	**362**
	7339	A	31	30	33	94
		B	31	30	31	92
		C	30	30	31	91
		D	30	30	30	90
			122	**120**	**125**	**367**
	8057	A	30	30	30	90
		B	30	30	30	90
		C	30	30	30	90
		D	30	30	30	90
			120	**120**	**120**	**360**
Old Totals			***363***	***361***	***365***	***1089***
New	8274	A	32	32	32	84
		B	32	84	84	258
		C	58	41	41	124
		D	80	80	80	235
			202	**237**	**237**	**676**
	8316	A	33	25	26	84
		B	84	84	90	258
		C	42	41	41	124
		D	80	80	75	235
			239	**230**	**232**	**701**
	9007	A	38	25	23	86
		B	76	76	76	228
		C	35	31	34	100
		D	84	85	89	258
			233	**217**	**222**	**672**
New Totals			***674***	***684***	***691***	***2049***
N total			***1037***	***1045***	***1056***	**3138**

Table 2: Means and Standard Deviations of Parameters

Level	Parameter	Mean	Standard Deviation
Level 1	WBC Count	3.075	0.099
	Neut %	49.170	3.817
	Lymph %	39.921	3.118
	Mono %	6.035	1.549
	Eos %	4.873	0.842
	Baso %	0.002	0.015
	RBC Count	2.909	0.039
	HGB	7.016	0.128
	HCT	22.471	0.422
	MCV	77.265	1.667
	MCH	24.122	0.608
	MCHC	31.225	0.630
	RDW	18.305	1.653
	PLT	78.472	2.912
	MPV	9.485	0.498
Level 2	WBC Count	6.796	0.393
	Neut %	55.743	4.520
	Lymph %	29.080	3.145
	Mono %	8.949	1.664
	Eos %	6.223	0.959
	Baso %	0.005	0.022
	RBC Count	4.285	0.064
	HGB	11.526	0.250
	HCT	36.482	0.956
	MCV	85.143	2.000
	MCH	26.904	0.572
	MCHC	31.601	0.478
	RDW	16.185	0.927
	PLT	221.146	8.001
	MPV	9.042	0.367
Level 3	WBC Count	16.197	0.592
	Neut %	64.778	3.737
	Lymph %	17.317	1.536
	Mono %	12.313	1.989
	Eos %	5.587	0.881
	Baso %	0.004	0.020
	RBC Count	5.221	0.077
	HGB	16.202	0.244
	HCT	50.152	0.813
	MCV	96.058	1.355
	MCH	31.032	0.442
	MCHC	32.309	0.463
	RDW	14.103	0.303
	PLT	601.041	22.707
	MPV	8.972	0.344

Table 3: One-Sample TOST Results for White Blood Cell Count, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	2.986	2.686 - 3.286	8274	A	3.209	<0.001	Yes
				B	3.125	<0.001	Yes
				C	3.164	<0.001	Yes
				D	3.225	<0.001	Yes
			8316	A	3.100	<0.001	Yes
				B	3.013	<0.001	Yes
				C	3.038	<0.001	Yes
				D	3.106	<0.001	Yes
			9007	A	3.163	<0.001	Yes
				B	3.072	<0.001	Yes
				C	3.094	<0.001	Yes
				D	3.176	<0.001	Yes
2	6.291	5.591 - 6.991	8274	A	7.081	1.000	**No**
				B	7.125	1.000	**No**
				C	7.173	1.000	**No**
				D	7.128	1.000	**No**
			8316	A	6.932	0.0043	Yes
				B	6.936	<0.001	Yes
				C	7.037	0.9986	**No**
				D	6.906	<0.001	Yes
			9007	A	7.108	0.999	**No**
				B	7.111	1.000	**No**
				C	7.227	1.000	**No**
				D	7.082	1.000	**No**
3	15.967	13.467 - 18.467	8274	A	16.344	<0.001	Yes
				B	16.392	<0.001	Yes
				C	16.683	<0.001	Yes
				D	16.568	<0.001	Yes
			8316	A	16.116	<0.001	Yes
				B	16.239	<0.001	Yes
				C	16.288	<0.001	Yes
				D	16.349	<0.001	Yes
			9007	A	16.096	<0.001	Yes
				B	16.033	<0.001	Yes
				C	16.571	<0.001	Yes
				D	16.179	<0.001	Yes

Table 4: One-Sample TOST Results for Neutrophil Percentage, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	49.836	44.836 - 54.836	8274	A	47.816	<0.001	Yes
				B	48.586	<0.001	Yes
				C	48.169	<0.001	Yes
				D	48.134	<0.001	Yes
			8316	A	45.206	0.026	Yes
				B	46.074	<0.001	Yes
				C	45.778	<0.001	Yes
				D	45.661	<0.001	Yes
			9007	A	52.111	<0.001	Yes
				B	52.980	<0.001	Yes
				C	52.441	<0.001	Yes
				D	52.175	<0.001	Yes
2	54.573	49.573 - 59.573	8274	A	55.506	<0.001	Yes
				B	56.232	<0.001	Yes
				C	55.385	<0.001	Yes
				D	55.818	<0.001	Yes
			8316	A	53.892	<0.001	Yes
				B	54.752	<0.001	Yes
				C	53.539	<0.001	Yes
				D	54.371	<0.001	Yes
			9007	A	58.952	<0.001	Yes
				B	59.986	1.000	**No**
				C	58.266	<0.001	Yes
				D	58.687	<0.001	Yes
3	62.577	57.577 - 67.577	8274	A	65.644	<0.001	Yes
				B	66.277	<0.001	Yes
				C	66.127	<0.001	Yes
				D	65.878	<0.001	Yes
			8316	A	64.788	<0.001	Yes
				B	66.043	<0.001	Yes
				C	65.220	<0.001	Yes
				D	65.232	<0.001	Yes
			9007	A	65.857	<0.001	Yes
				B	67.067	<0.001	Yes
				C	66.118	<0.001	Yes
				D	65.806	<0.001	Yes

Table 5: One-Sample TOST Results for Lymphocyte Percentage, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	40.655	34.655 - 46.655	8274	A	41.213	<0.001	Yes
				B	40.410	<0.001	Yes
				C	40.797	<0.001	Yes
				D	40.839	<0.001	Yes
			8316	A	43.221	<0.001	Yes
				B	42.510	<0.001	Yes
				C	42.705	<0.001	Yes
				D	42.948	<0.001	Yes
			9007	A	35.305	<0.001	Yes
				B	34.609	0.733	**No**
				C	35.112	0.002	Yes
				D	35.359	<0.001	Yes
2	29.213	24.213 - 34.213	8274	A	29.781	<0.001	Yes
				B	29.032	<0.001	Yes
				C	30.208	<0.001	Yes
				D	29.605	<0.001	Yes
			8316	A	32.304	<0.001	Yes
				B	31.605	<0.001	Yes
				C	32.805	<0.001	Yes
				D	32.111	<0.001	Yes
			9007	A	25.400	<0.001	Yes
				B	24.274	0.215	**No**
				C	25.921	<0.001	Yes
				D	25.644	<0.001	Yes
3	18.000	13.000 - 23.000	8274	A	17.247	<0.001	Yes
				B	16.435	<0.001	Yes
				C	17.176	<0.001	Yes
				D	17.321	<0.001	Yes
			8316	A	18.324	<0.001	Yes
				B	17.333	<0.001	Yes
				C	18.405	<0.001	Yes
				D	18.521	<0.001	Yes
			9007	A	16.022	<0.001	Yes
				B	14.929	<0.001	Yes
				C	16.121	<0.001	Yes
				D	16.480	<0.001	Yes

Table 6: One-Sample TOST Results for Monocyte Percentage, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	4.570	1.570 - 7.570	8274	A	6.841	<0.001	Yes
				B	6.848	<0.001	Yes
				C	6.683	<0.001	Yes
				D	6.740	<0.001	Yes
			8316	A	6.485	0.011	Yes
				B	6.699	<0.001	Yes
				C	6.580	<0.001	Yes
				D	6.531	<0.001	Yes
			9007	A	7.282	<0.001	Yes
				B	7.203	<0.001	Yes
				C	7.044	<0.001	Yes
				D	7.063	<0.001	Yes
2	9.294	6.294 - 12.294	8274	A	8.806	<0.001	Yes
				B	8.956	<0.001	Yes
				C	8.475	<0.001	Yes
				D	8.639	<0.001	Yes
			8316	A	8.416	<0.001	Yes
				B	8.524	<0.001	Yes
				C	8.000	<0.001	Yes
				D	8.254	<0.001	Yes
			9007	A	9.376	<0.001	Yes
				B	9.404	<0.001	Yes
				C	9.145	<0.001	Yes
				D	9.140	<0.001	Yes
3	13.331	10.331 - 16.331	8274	A	11.872	<0.001	Yes
				B	12.274	<0.001	Yes
				C	11.305	<0.001	Yes
				D	11.501	<0.001	Yes
			8316	A	11.652	<0.001	Yes
				B	11.817	<0.001	Yes
				C	10.929	<0.001	Yes
				D	11.141	<0.001	Yes
			9007	A	12.444	<0.001	Yes
				B	12.488	<0.001	Yes
				C	11.856	<0.001	Yes
				D	11.857	<0.001	Yes

Table 7: One-Sample TOST Results for Eosinophil Percentage, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	4.938	1.938 - 7.938	8274	A	4.131	<0.001	Yes
				B	4.155	<0.001	Yes
				C	4.352	<0.001	Yes
				D	4.285	<0.001	Yes
			8316	A	5.088	<0.001	Yes
				B	4.713	<0.001	Yes
				C	4.930	<0.001	Yes
				D	4.855	<0.001	Yes
			9007	A	5.303	<0.001	Yes
				B	5.203	<0.001	Yes
				C	5.400	<0.001	Yes
				D	5.400	<0.001	Yes
2	6.914	2.914 - 10.914	8274	A	5.906	<0.001	Yes
				B	5.780	<0.001	Yes
				C	5.933	<0.001	Yes
				D	5.938	<0.001	Yes
			8316	A	5.376	<0.001	Yes
				B	5.112	<0.001	Yes
				C	5.644	<0.001	Yes
				D	5.255	<0.001	Yes
			9007	A	6.268	<0.001	Yes
				B	6.333	<0.001	Yes
				C	6.662	<0.001	Yes
				D	6.526	<0.001	Yes
3	6.093	2.093 - 10.093	8274	A	5.238	<0.001	Yes
				B	5.013	<0.001	Yes
				C	5.393	<0.001	Yes
				D	5.300	<0.001	Yes
			8316	A	5.220	<0.001	Yes
				B	4.793	<0.001	Yes
				C	5.424	<0.001	Yes
				D	5.101	<0.001	Yes
			9007	A	5.674	<0.001	Yes
				B	5.509	<0.001	Yes
				C	5.897	<0.001	Yes
				D	5.856	<0.001	Yes

Table 8: One-Sample TOST Results for Basophil Percentage, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	0.001	-0.499 - 0.501	8274	A	0.000	<0.001	Yes
				B	0.000	<0.001	Yes
				C	0.000	<0.001	Yes
				D	0.003	<0.001	Yes
			8316	A	0.000	<0.001	Yes
				B	0.000	<0.001	Yes
				C	0.000	<0.001	Yes
				D	0.000	<0.001	Yes
			9007	A	0.000	<0.001	Yes
				B	0.005	<0.001	Yes
				C	0.003	<0.001	Yes
				D	0.002	<0.001	Yes
2	0.007	-0.493 - 0.507	8274	A	0.000	<0.001	Yes
				B	0.000	<0.001	Yes
				C	0.000	<0.001	Yes
				D	0.000	<0.001	Yes
			8316	A	0.000	<0.001	Yes
				B	0.000	<0.001	Yes
				C	0.000	<0.001	Yes
				D	0.000	<0.001	Yes
			9007	A	0.004	<0.001	Yes
				B	0.000	<0.001	Yes
				C	0.006	<0.001	Yes
				D	0.000	<0.001	Yes
3	0.002	-0.498 - 0.502	8274	A	0.001	<0.001	Yes
				B	0.000	<0.001	Yes
				C	0.000	<0.001	Yes
				D	0.000	<0.001	Yes
			8316	A	0.000	<0.001	Yes
				B	0.000	<0.001	Yes
				C	0.000	<0.001	Yes
				D	0.000	<0.001	Yes
			9007	A	0.004	<0.001	Yes
				B	0.000	<0.001	Yes
				C	0.000	<0.001	Yes
				D	0.001	<0.001	Yes

Table 9: One-Sample TOST Results for Red Blood Cell Count, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	2.938	2.788 - 3.088	8274	A	2.861	<0.001	Yes
				B	2.863	<0.001	Yes
				C	2.896	<0.001	Yes
				D	2.909	<0.001	Yes
			8316	A	2.869	<0.001	Yes
				B	2.881	<0.001	Yes
				C	2.910	<0.001	Yes
				D	2.907	<0.001	Yes
			9007	A	2.876	<0.001	Yes
				B	2.880	<0.001	Yes
				C	2.905	<0.001	Yes
				D	2.913	<0.001	Yes
2	4.253	4.053 - 4.453	8274	A	4.294	<0.001	Yes
				B	4.336	<0.001	Yes
				C	4.314	<0.001	Yes
				D	4.354	<0.001	Yes
			8316	A	4.246	<0.001	Yes
				B	4.283	<0.001	Yes
				C	4.271	<0.001	Yes
				D	4.291	<0.001	Yes
			9007	A	4.254	<0.001	Yes
				B	4.301	<0.001	Yes
				C	4.277	<0.001	Yes
				D	4.298	<0.001	Yes
3	5.247	4.947 - 5.547	8274	A	5.277	<0.001	Yes
				B	5.306	<0.001	Yes
				C	5.261	<0.001	Yes
				D	5.301	<0.001	Yes
			8316	A	5.170	<0.001	Yes
				B	5.159	<0.001	Yes
				C	5.137	<0.001	Yes
				D	5.143	<0.001	Yes
			9007	A	5.190	<0.001	Yes
				B	5.182	<0.001	Yes
				C	5.171	<0.001	Yes
				D	5.169	<0.001	Yes

Table 10: One-Sample TOST Results for Hemoglobin, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	6.896	6.496 - 7.296	8274	A	6.994	<0.001	Yes
				B	7.134	<0.001	Yes
				C	7.010	<0.001	Yes
				D	7.058	<0.001	Yes
			8316	A	7.021	<0.001	Yes
				B	7.160	<0.001	Yes
				C	7.029	<0.001	Yes
				D	7.084	<0.001	Yes
			9007	A	7.024	<0.001	Yes
				B	7.164	<0.001	Yes
				C	7.054	<0.001	Yes
				D	7.093	<0.001	Yes
2	11.366	10.766 - 11.966	8274	A	11.213	<0.001	Yes
				B	11.464	<0.001	Yes
				C	11.198	<0.001	Yes
				D	11.366	<0.001	Yes
			8316	A	11.564	<0.001	Yes
				B	11.769	<0.001	Yes
				C	11.544	<0.001	Yes
				D	11.709	<0.001	Yes
			9007	A	11.684	<0.001	Yes
				B	11.905	<0.001	Yes
				C	11.668	<0.001	Yes
				D	11.825	<0.001	Yes
3	16.219	15.519 - 16.919	8274	A	16.253	<0.001	Yes
				B	16.590	<0.001	Yes
				C	16.195	<0.001	Yes
				D	16.398	<0.001	Yes
			8316	A	15.958	<0.001	Yes
				B	16.170	<0.001	Yes
				C	15.895	<0.001	Yes
				D	16.036	<0.001	Yes
			9007	A	16.013	<0.001	Yes
				B	16.230	<0.001	Yes
				C	15.932	<0.001	Yes
				D	16.083	<0.001	Yes

Table 11: One-Sample TOST Results for Hematocrit, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	22.155	20.655 - 23.655	8274	A	22.178	<0.001	Yes
				B	22.252	<0.001	Yes
				C	22.548	<0.001	Yes
				D	22.550	<0.001	Yes
			8316	A	22.497	<0.001	Yes
				B	22.740	<0.001	Yes
				C	22.740	<0.001	Yes
				D	22.778	<0.001	Yes
			9007	A	22.616	<0.001	Yes
				B	22.653	<0.001	Yes
				C	22.729	<0.001	Yes
				D	22.857	<0.001	Yes
2	35.929	33.529 - 38.329	8274	A	35.438	<0.001	Yes
				B	35.938	<0.001	Yes
				C	35.744	<0.001	Yes
				D	35.946	<0.001	Yes
			8316	A	36.736	<0.001	Yes
				B	37.360	<0.001	Yes
				C	36.934	<0.001	Yes
				D	37.154	<0.001	Yes
			9007	A	37.092	<0.001	Yes
				B	37.550	<0.001	Yes
				C	37.126	<0.001	Yes
				D	37.459	<0.001	Yes
3	49.887	46.887 - 52.887	8274	A	50.484	<0.001	Yes
				B	50.961	<0.001	Yes
				C	50.571	<0.001	Yes
				D	50.730	<0.001	Yes
			8316	A	50.219	<0.001	Yes
				B	50.406	<0.001	Yes
				C	49.768	<0.001	Yes
				D	49.897	<0.001	Yes
			9007	A	50.165	<0.001	Yes
				B	50.138	<0.001	Yes
				C	49.729	<0.001	Yes
				D	49.922	<0.001	Yes

Table 12: One-Sample TOST Results for Mean Cell Volume, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	75.410	69.910 - 80.910	8274	A	77.525	<0.001	Yes
				B	77.721	<0.001	Yes
				C	77.859	<0.001	Yes
				D	77.505	<0.001	Yes
			8316	A	78.406	<0.001	Yes
				B	78.924	<0.001	Yes
				C	78.171	<0.001	Yes
				D	78.378	<0.001	Yes
			9007	A	78.674	<0.001	Yes
				B	78.639	<0.001	Yes
				C	78.229	<0.001	Yes
				D	78.481	<0.001	Yes
2	84.479	78.979 - 89.979	8274	A	82.519	<0.001	Yes
				B	82.880	<0.001	Yes
				C	82.824	<0.001	Yes
				D	82.563	<0.001	Yes
			8316	A	86.516	<0.001	Yes
				B	87.217	<0.001	Yes
				C	86.468	<0.001	Yes
				D	86.586	<0.001	Yes
			9007	A	87.184	<0.001	Yes
				B	87.304	<0.001	Yes
				C	86.784	<0.001	Yes
				D	87.156	<0.001	Yes
3	95.072	89.572 - 100.572	8274	A	95.663	<0.001	Yes
				B	96.051	<0.001	Yes
				C	96.137	<0.001	Yes
				D	95.681	<0.001	Yes
			8316	A	97.127	<0.001	Yes
				B	97.697	<0.001	Yes
				C	96.878	<0.001	Yes
				D	97.020	<0.001	Yes
			9007	A	96.678	<0.001	Yes
				B	96.759	<0.001	Yes
				C	96.191	<0.001	Yes
				D	96.578	<0.001	Yes

Table 13: One-Sample TOST Results for Mean Cell Hemoglobin, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	23.465	21.765 - 25.165	8274	A	24.419	<0.001	Yes
				B	24.941	0.038	Yes
				C	24.205	<0.001	Yes
				D	24.261	<0.001	Yes
			8316	A	24.467	<0.001	Yes
				B	24.857	<0.001	Yes
				C	24.160	<0.001	Yes
				D	24.375	<0.001	Yes
			9007	A	24.429	<0.001	Yes
				B	24.878	<0.001	Yes
				C	24.280	<0.001	Yes
				D	24.352	<0.001	Yes
2	26.733	25.033 - 28.433	8274	A	26.113	<0.001	Yes
				B	26.442	<0.001	Yes
				C	25.934	<0.001	Yes
				D	26.108	<0.001	Yes
			8316	A	27.244	<0.001	Yes
				B	27.471	<0.001	Yes
				C	27.029	<0.001	Yes
				D	27.285	<0.001	Yes
			9007	A	27.464	<0.001	Yes
				B	27.682	<0.001	Yes
				C	27.284	<0.001	Yes
				D	27.520	<0.001	Yes
3	30.913	29.213 - 32.613	8274	A	30.797	<0.001	Yes
				B	31.270	<0.001	Yes
				C	30.790	<0.001	Yes
				D	30.931	<0.001	Yes
			8316	A	30.873	<0.001	Yes
				B	31.348	<0.001	Yes
				C	30.937	<0.001	Yes
				D	31.187	<0.001	Yes
			9007	A	30.857	<0.001	Yes
				B	31.314	<0.001	Yes
				C	30.809	<0.001	Yes
				D	31.112	<0.001	Yes

Table 14: One-Sample TOST Results for Mean Cell Hemoglobin Concentration, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	31.133	28.633 - 33.633	8274	A	31.494	<0.001	Yes
				B	32.097	<0.001	Yes
				C	31.090	<0.001	Yes
				D	31.304	<0.001	Yes
			8316	A	31.212	<0.001	Yes
				B	31.490	<0.001	Yes
				C	30.919	<0.001	Yes
				D	31.098	<0.001	Yes
			9007	A	31.042	<0.001	Yes
				B	31.636	<0.001	Yes
				C	31.043	<0.001	Yes
				D	31.032	<0.001	Yes
2	31.648	29.148 - 34.148	8274	A	31.647	<0.001	Yes
				B	31.902	<0.001	Yes
				C	31.327	<0.001	Yes
				D	31.619	<0.001	Yes
			8316	A	31.488	<0.001	Yes
				B	31.504	<0.001	Yes
				C	31.259	<0.001	Yes
				D	31.523	<0.001	Yes
			9007	A	31.500	<0.001	Yes
				B	31.707	<0.001	Yes
				C	31.439	<0.001	Yes
				D	31.572	<0.001	Yes
3	32.515	30.015 - 35.015	8274	A	32.203	<0.001	Yes
				B	32.561	<0.001	Yes
				C	32.017	<0.001	Yes
				D	32.326	<0.001	Yes
			8316	A	31.796	<0.001	Yes
				B	32.087	<0.001	Yes
				C	31.937	<0.001	Yes
				D	32.140	<0.001	Yes
			9007	A	31.922	<0.001	Yes
				B	32.374	<0.001	Yes
				C	32.035	<0.001	Yes
				D	32.217	<0.001	Yes

Table 15: One-Sample TOST Results for Red Blood Cell Size Distribution, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	20.321	18.321 - 22.321	8274	A	16.522	1.000	**No**
				B	16.550	1.000	**No**
				C	16.486	1.000	**No**
				D	16.570	1.000	**No**
			8316	A	18.306	0.601	**No**
				B	18.357	0.091	**No**
				C	18.374	0.089	**No**
				D	18.266	0.967	**No**
			9007	A	16.624	1.000	**No**
				B	16.722	1.000	**No**
				C	16.703	1.000	**No**
				D	16.663	1.000	**No**
2	15.720	13.720 - 17.720	8274	A	17.722	0.516	**No**
				B	17.773	0.987	**No**
				C	17.732	0.627	**No**
				D	17.690	0.129	**No**
			8316	A	16.116	<0.001	Yes
				B	16.135	<0.001	Yes
				C	16.154	<0.001	Yes
				D	16.078	<0.001	Yes
			9007	A	15.312	<0.001	Yes
				B	15.408	<0.001	Yes
				C	15.323	<0.001	Yes
				D	15.309	<0.001	Yes
3	14.255	12.255 - 16.255	8274	A	13.741	<0.001	Yes
				B	13.764	<0.001	Yes
				C	13.749	<0.001	Yes
				D	13.745	<0.001	Yes
			8316	A	14.312	<0.001	Yes
				B	14.361	<0.001	Yes
				C	14.341	<0.001	Yes
				D	14.277	<0.001	Yes
			9007	A	14.013	<0.001	Yes
				B	14.009	<0.001	Yes
				C	13.971	<0.001	Yes
				D	13.991	<0.001	Yes

Table 16: One-Sample TOST Results for Platelet Count, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	76.353	61.353 - 91.353	8274	A	76.438	<0.001	Yes
				B	77.233	<0.001	Yes
				C	80.345	<0.001	Yes
				D	79.975	<0.001	Yes
			8316	A	76.242	<0.001	Yes
				B	77.488	<0.001	Yes
				C	79.905	<0.001	Yes
				D	80.888	<0.001	Yes
			9007	A	79.132	<0.001	Yes
				B	79.368	<0.001	Yes
				C	81.457	<0.001	Yes
				D	82.619	<0.001	Yes
2	223.371	193.371 - 253.371	8274	A	217.813	<0.001	Yes
				B	221.179	<0.001	Yes
				C	225.098	<0.001	Yes
				D	224.800	<0.001	Yes
			8316	A	208.120	0.031	Yes
				B	211.976	<0.001	Yes
				C	216.878	<0.001	Yes
				D	216.938	<0.001	Yes
			9007	A	218.320	<0.001	Yes
				B	221.224	<0.001	Yes
				C	224.710	<0.001	Yes
				D	225.953	<0.001	Yes
3	622.124	562.124 - 682.124	8274	A	581.313	<0.001	Yes
				B	591.096	<0.001	Yes
				C	596.439	<0.001	Yes
				D	597.750	<0.001	Yes
			8316	A	565.731	<0.001	Yes
				B	575.300	<0.001	Yes
				C	584.488	<0.001	Yes
				D	587.080	<0.001	Yes
			9007	A	586.609	<0.001	Yes
				B	592.645	<0.001	Yes
				C	600.706	<0.001	Yes
				D	603.022	<0.001	Yes

Table 17: One-Sample TOST Results for Mean Platelet Volume, Stratified by Lot and Instrument

Level	Target	Equivalence Range	Lot	Instrument	Mean	p=value	Equivalent?
1	10.057	8.957 - 11.157	8274	A	9.316	<0.001	Yes
				B	9.413	<0.001	Yes
				C	9.634	<0.001	Yes
				D	9.151	<0.001	Yes
			8316	A	9.064	1.000	**No**
				B	9.093	<0.001	Yes
				C	9.336	<0.001	Yes
				D	8.888	<0.001	Yes
			9007	A	9.126	<0.001	Yes
				B	9.199	<0.001	Yes
				C	9.426	<0.001	Yes
				D	8.962	0.315	**No**
2	9.037	7.937 - 10.137	8274	A	9.188	<0.001	Yes
				B	9.221	<0.001	Yes
				C	9.488	<0.001	Yes
				D	8.997	<0.001	Yes
			8316	A	8.884	<0.001	Yes
				B	8.912	<0.001	Yes
				C	9.159	<0.001	Yes
				D	8.706	<0.001	Yes
			9007	A	9.100	<0.001	Yes
				B	9.129	<0.001	Yes
				C	9.352	<0.001	Yes
				D	8.887	<0.001	Yes
3	8.961	7.861 - 10.061	8274	A	9.150	<0.001	Yes
				B	9.168	<0.001	Yes
				C	9.424	<0.001	Yes
				D	8.941	<0.001	Yes
			8316	A	8.892	<0.001	Yes
				B	8.864	<0.001	Yes
				C	9.112	<0.001	Yes
				D	8.648	<0.001	Yes
			9007	A	8.996	<0.001	Yes
				B	9.016	<0.001	Yes
				C	9.276	<0.001	Yes
				D	8.764	<0.001	Yes

Table 18: Two-Sample TOST Results for All Parameters by Level (Lots and Instruments Combined)

Parameter	Level	Equivalence Bounds	Mean Difference (95% CI)	p-value	Equivalent?
WBC Count	1	-0.3, 0.3	0.1367 (0.1288, 0.1447)	<0.0001	Yes
	2	**-0.7, 0.7**	**0.7712 (0.7561, 0.7864)**	**1.0000**	**No**
	3	-2.5, 2.5	0.3515 (0.2908, 0.4122)	<0.0001	Yes
Neutrophil %	1	-5.0, 5.0	-1.0285 (-1.4358, -0.6213)	<0.0001	Yes
	2	-5.0, 5.0	1.7904 (1.3139, 2.2670)	<0.0001	Yes
	3	-5.0, 5.0	3.3585 (2.9977, 3.7193)	<0.0001	Yes
Lymphocyte %	1	-6.0, 6.0	-1.1341 (-1.4644, -0.8038)	<0.0001	Yes
	2	-5.0, 5.0	-0.2032 (-0.5407, 0.1342)	<0.0001	Yes
	3	-5.0, 5.0	-1.0375 (-1.1929, -0.8822)	<0.0001	Yes
Monocyte %	1	-3.0, 3.0	2.2611 (2.1418, 2.3805)	<0.0001	Yes
	2	-3.0, 3.0	-0.5267 (-0.7033, -0.3501)	<0.0001	Yes
	3	-3.0, 3.0	-1.5528 (-1.75000, -1.3555)	<0.0001	Yes
Eosinophil %	1	-3.0, 3.0	-0.1006 (-0.1909, -0.0102)	<0.0001	Yes
	2	-4.0, 4.0	-1.0580 (-1.1456, -0.9704)	<0.0001	Yes
	3	-4.0, 4.0	-0.7720 (-0.8575, -0.6864)	<0.0001	Yes
Basophil %	1	-0.5, 0.5	0.0020 (0.0004, 0.0036)	<0.0001	Yes
	2	-0.5, 0.5	-0.0025 (-0.0049, 0.0001)	<0.0001	Yes
	3	-0.5, 0.5	0.0037 (0.0016, 0.0059)	<0.0001	Yes
RBC Count	1	-0.15, 0.15	-0.0453 (-0.0488, -0.0418)	<0.0001	Yes
	2	-0.2, 0.2	0.0491, (0.0427, 0.0554)	<0.0001	Yes
	3	-0.3, 0.3	-0.0393 (-0.0473, 0.0314)	<0.0001	Yes
Hemoglobin	1	-0.4, 0.4	0.1849 (0.1750, 0.1948)	<0.0001	Yes
	2	-0.6, 0.6	0.2446 (0.2208, 0.2684)	<0.0001	Yes
	3	-0.7, 0.7	-0.0262 (-0.0521, -0.0002)	<0.0001	Yes
Hematocrit	1	-1.5, 1.5	0.4875 (0.4497, 0.5253)	<0.0001	Yes
	2	-2.4, 2.4	0.8443 (0.7514, 0.9373)	<0.0001	Yes
	3	-3.0, 3.0	0.4041 (0.3198, 0.4883)	<0.0001	Yes
Mean Cell Volume	1	-5.5, 5.5	2.8579 (2.7551, 2.9607)	<0.0001	Yes
	2	-5.5, 5.5	1.0149 (0.8070, 1.2228)	<0.0001	Yes
	3	-5.5, 5.5	1.5062 (1.3835, 1.6288)	<0.0001	Yes
Mean Cell Hemoglobin	1	-1.7, 1.7	1.0116 (0.9721, 1.0512)	<0.0001	Yes
	2	-1.7, 1.7	0.2605 (0.2007, 0.3203)	<0.0001	Yes
	3	-1.7, 1.7	0.1820 (0.1358, 0.2282)	<0.0001	Yes
Mean Cell Hemoglobin Concentration	1	-2.5, 2.5	0.1406 (0.0734, 0.2078)	<0.0001	Yes
	2	-2.5, 2.5	-0.0706 (-0.1217, -0.0195)	<0.0001	Yes
	3	-2.5, 2.5	-0.3158 (-0.3625, -0.2691)	<0.0001	Yes
Red Blood Cell Size Distribution	**1**	**-2.0, 2.0**	**-3.1018 (-3.1806, -3.0230)**	**1.0000**	**No**
	2	-2.0, 2.0	0.7119 (0.6194, 0.8044)	<0.0001	Yes
	3	-2.0, 2.0	-0.2321 (-0.2621, -0.2021)	<0.0001	Yes
Platelet Count	1	-15.0, 15.0	3.2649 (3.0010, 3.5288)	<0.0001	Yes
	2	-30.0, 30.0	-3.3990 (-4.2386, -2.5594)	<0.0001	Yes
	3	-60.0, 60.0	-32.2048 (-33.9931, -30.4164)	<0.0001	Yes
Mean Platelet Volume	1	-1.1, 1.1	-0.8814 (-0.9099, -0.8529)	<0.0001	Yes
	2	-1.1, 1.1	0.0076 (-0.0317, 0.0469)	<0.0001	Yes
	3	-1.1, 1.1	0.0163 (-0.0204, 0.0530)	<0.0001	Yes

Table 19: One-Sample TOST Results for All Parameters by Level, Using Old Formulation Mean and Equivalence Range as Targets (Lots and Instruments Combined)

Parameter	Level	Target Value	Equivalence Range	Mean (95% CI)	p-value	Equivalent?
WBC Count	1	2.986 ± 0.3	2.686 - 3.286	3.1224 (3.1172, 3.1276)	<0.0001	Yes
	2	**6.291 ± 0.7**	**5.591 - 6.991**	**7.0618 (7.0529, 7.0707)**	**1.0000**	**No**
	3	15.967 ± 2.5	13.467 - 18.467	16.3184 (16.2851, 16.3517)	<0.0001	Yes
Neutrophil %	1	49.836 ± 5.0	44.836 - 54.836	48.8074 (48.6190, 48.9958)	<0.0001	Yes
	2	54.573 ± 5.0	49.573 - 59.573	56.3629 (56.2259, 56.5000)	<0.0001	Yes
	3	62.577 ± 5.0	57.577 - 67.577	65.9359 (65.8799, 65.9919)	<0.0001	Yes
Lymphocyte %	1	40.655 ± 6.0	34.655 - 46.655	39.5209 (39.3024, 39.7394)	<0.0001	Yes
	2	29.213 ± 5.0	24.213 - 34.213	29.0093 (28.8241, 29.1945)	<0.0001	Yes
	3	18.000 ± 5.0	13.000 - 23.000	16.9594 (16.8865, 17.0323)	<0.0001	Yes
Monocyte %	1	4.570 ± 3.0	1.570 - 7.570	6.8316 (6.8031, 6.8600)	<0.0001	Yes
	2	9.294 ± 3.0	6.294 - 12.294	8.7669 (8.7293, 8.8045)	<0.0001	Yes
	3	13.331 ± 3.0	10.331 - 16.331	11.7778 (11.7354, 11.8202)	<0.0001	Yes
Eosinophil %	1	4.938 ± 3.0	1.938 - 7.938	4.8373 (4.7965, 4.8781)	<0.0001	Yes
	2	6.914 ± 4.0	2.914 - 10.914	5.8565 (5.8129, 5.9000)	<0.0001	Yes
	3	6.093 ± 4.0	2.093 - 10.093	5.3212 (5.2850, 5.3573)	<0.0001	Yes
Basophil %	1	0.001 ± 0.5	-0.499 - 0.501	0.0028 (0.0017, 0.0040)	<0.0001	Yes
	2	0.007 ± 0.5	-0.493 - 0.507	0.0044 (0.0031, 0.0057)	<0.0001	Yes
	3	0.002 ± 0.5	-0.498 - 0.502	0.0057 (0.0042, 0.0071)	<0.0001	Yes
RBC Count	1	2.938 ± 0.15	2.788 - 3.088	2.8929 (2.8910, 2.8948)	<0.0001	Yes
	2	4.253 ± 0.2	4.053 - 4.453	4.3017 (4.2989, 4.3045)	<0.0001	Yes
	3	5.247 ± 0.3	4.947 - 5.547	5.2076 (5.2026, 5.2126)	<0.0001	Yes
Hemoglobin	1	6.896 ± 0.4	6.496 - 7.296	7.0805 (7.0754, 7.0856)	<0.0001	Yes
	2	11.366 ± 0.6	10.766 - 11.966	11.6105 (11.5959, 11.6252)	<0.0001	Yes
	3	16.219 ± 0.7	15.519 - 16.919	16.1925 (16.1779, 16.2071)	<0.0001	Yes
Hematocrit	1	22.155 ± 1.5	20.655 - 23.655	22.6426 (22.6208, 22.6645)	<0.0001	Yes
	2	35.929 ± 2.4	33.529 - 38.329	36.7737 (36.7228, 36.8245)	<0.0001	Yes
	3	49.887 ± 3.0	46.887 - 52.887	50.2912 (50.2489, 50.3335)	<0.0001	Yes
Mean Cell Volume	1	75.410 ± 5.5	69.910 - 80.910	78.2681 (78.2100, 78.3262)	<0.0001	Yes
	2	84.479 ± 5.5	78.979 - 89.979	85.4939 (85.3574, 85.6303)	<0.0001	Yes
	3	95.072 ± 5.5	89.572 - 100.572	96.5779 (96.5164, 96.6393)	<0.0001	Yes
Mean Cell Hemoglobin	1	23.465 ± 1.7	21.765 - 25.165	24.4769 (24.4530, 24.5008)	<0.0001	Yes
	2	26.733 ± 1.7	25.033 - 28.433	26.9937 (26.9521, 27.0353)	<0.0001	Yes
	3	30.913 ± 1.7	29.213 - 32.613	31.0949 (31.0725, 31.1174)	<0.0001	Yes
Mean Cell Hemoglobin Concentration	1	31.133 ± 2.5	28.633 - 33.633	31.2739 (31.2398, 31.3081)	<0.0001	Yes
	2	31.648 ± 2.5	29.148 - 34.148	31.5770 (31.5525, 31.6016)	<0.0001	Yes
	3	32.515 ± 2.5	30.015 - 35.015	32.1996 (32.1732, 32.2259)	<0.0001	Yes
Red Blood Cell Size Distribution	**1**	**20.321 ± 2.0**	**18.321 - 22.321**	**17.2197 (17.1656, 17.2738)**	**1.0000**	**No**
	2	15.720 ± 2.0	13.720 - 17.720	16.4313 (16.3671, 16.4955)	<0.0001	Yes
	3	14.255 ± 2.0	12.255 - 16.255	14.0226 (14.0045, 14.0407)	<0.0001	Yes
Platelet Count	1	76.353 ± 15.0	61.353 - 91.353	79.6176 (79.4593, 79.7759)	<0.0001	Yes
	2	223.371 ± 30.0	193.371 - 253.371	220.000 (219.600, 220.400)	<0.0001	Yes
	3	622.124 ± 60.0	562.124 – 682.124	589.900 (589.000, 590.900)	<0.0001	Yes
Mean Platelet Volume	1	10.057 ± 1.1	8.957 - 11.157	9.1756 (9.1612, 9.1900)	<0.0001	Yes
	2	9.037 ± 1.1	7.937 - 10.137	9.0447 (9.0309, 9.0586)	<0.0001	Yes
	3	8.961 ± 1.1	7.861 - 10.061	8.9773 (8.9632, 8.9914)	<0.0001	Yes

Appendix B: Literature Cited

American Clinical Laboratory Association. (2014). *Importance of Clinical Lab Testing Highlighted During Medical Lab Professionals Week.*

Armijo-Olivo S. (2018). The importance of determining the clinical significance of research results in physical therapy clinical research. *Brazilian journal of physical therapy*, 22(3), 175–176. https://doi.org/10.1016/j.bjpt.2018.02.001

Bio-Rad Laboratories: Quality Systems Division. (2008). *Basic Lessons in Laboratory Quality Control*

Bonini, P., Plebani, M., Ceriotti, F., Rubboli, F. (2002). Errors in Laboratory Medicine. *Clinical Chemistry*, 48 (5), 691-698.

Centers for Disease Control and Prevention: Division of Laboratory Systems (DLS). (2018). *Strengthening Clinical Laboratories.*

Clinical Laboratory Improvement Amendments of 1988, 42 USC § 263a

Clinlab Navigator. (n.d.). Retrieved from http://www.clinlabnavigator.com/quality-control.html

Dixon, PM., Saint-Maurice, PF., Kim, Y., Hibbing, P., Bai, Y., Welk, GJ. (2018). A Primer on the Use of Equivalency Testing for Evaluating Measurement Agreement. *Medicine and science in sports and exercise*, 50(4), 837-845.

Ezzelle, J., Rodriguez-Chavez, I. R., Darden, J. M., Stirewalt, M., Kunwar, N., Hitchcock, R., Walter, T., & D'Souza, M. P. (2008). Guidelines on good clinical laboratory practice: bridging operations between research and clinical research laboratories. *Journal of pharmaceutical and biomedical analysis*, 46(1), 18–29. https://doi.org/10.1016/j.jpba.2007.10.010

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, 10(2), 486–489. https://doi.org/10.5812/ijem.3505

Ialongo, C. (2017). The logic of equivalence testing and its use in laboratory medicine. *Biochemia Medica*, 5–13. doi: 10.11613/bm.2017.001

Keohane, E. M., Walenga, J. M., & Otto, C. N. (2020). *Rodaks hematology: clinical principles and applications*. St. Louis, Missouri.: Elsevier.

Kotlarz VR. (1998). Tracing our roots: origins of clinical laboratory science. Clinical Laboratory Science: *Journal of the American Society for Medical Technology*. 11(1), 5-7

Lakens, D. (2017). Equivalence Tests. *Social Psychological and Personality Science*, 8(4), 355–362. doi: 10.1177/1948550617697177

Lauzon, C., & Caffo, B. (2009). Easy Multiplicity Control in Equivalence Testing Using Two One-sided Tests. *The American statistician*, 63(2), 147–154. doi:10.1198/tast.2009.0029

Li, X., Qiu, W., Morrow, J., DeMeo, D., Weiss, S., Fu, Y., Wang, X. (2015). A Comparative Study of Tests for Homogeneity of Variances with Application to DNA Methylation Data. *PLoS ONE*, 10(12). doi: 10.1371/journal.pone.0145295

Limentani, G. B., Ringo, M. C., Ye, F., Bergquist, M. L., & Mcsorley, E. O. (2005). Beyond the t-Test: Statistical Equivalence Testing. *Analytical Chemistry*, 77(11). doi: 10.1021/ac053390m

Lung, K. R., Gorko, M. A., Llewelyn, J., & Wiggins, N. (2003). Statistical method for the determination of equivalence of automated test procedures. *Journal of Automated Methods & Management in Chemistry*, 25(6), 123–127. doi: 10.1080/14639240410001703465

Medical Laboratory Observer. (2015). *Quality Control in Clinical Laboratory Samples*

Ngo, A., Gandhi, P., & Miller, W. G. (2016). Frequency that Laboratory Tests Influence Medical Decisions. *The Journal of Applied Laboratory Medicine: An AACC Publication*, 1(4), 410–414. doi: 10.1373/jalm.2016.021634

Njoroge, S. W., & Nichols, J. H. (2014). Risk management in the clinical laboratory. *Annals of laboratory medicine*, 34(4), 274–278. https://doi.org/10.3343/alm.2014.34.4.274

Schuirmann, D. J. (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. doi: 10.1007/bf01068419

Sheridan, C. (2020, February 19). Coronavirus and the race to distribute reliable diagnostics. Retrieved from https://www.nature.com/articles/d41587-020-00002-2

U.S. Food and Drug Administration. (2020). *Clinical Laboratory Improvement Amendments (CLIA).*

Voet, H. V. D., Goedhart, P. W., & Schmidt, K. (2017). Equivalence testing using existing reference data: An example with genetically modified and conventional crops in animal feeding studies. *Food and Chemical Toxicology*, 109, 472–485. doi: 10.1016/j.fct.2017.09.044

Wilson, M. L., Gradus, S., & Zimmerman, S. J. (2010). The role of local public health laboratories. *Public health reports* (Washington, D.C.: 1974), 125 Suppl 2(Suppl 2), 118–122. https://doi.org/10.1177/00333549101250S215

World Health Organization. (2009). *Good Clinical Laboratory Practice (GCLP)*