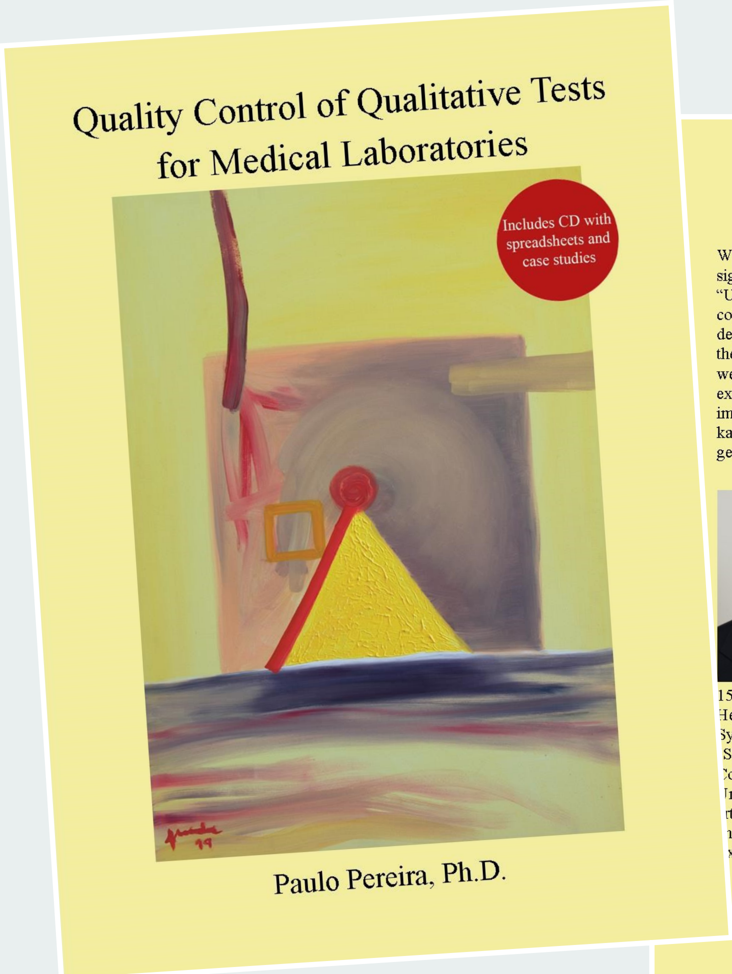
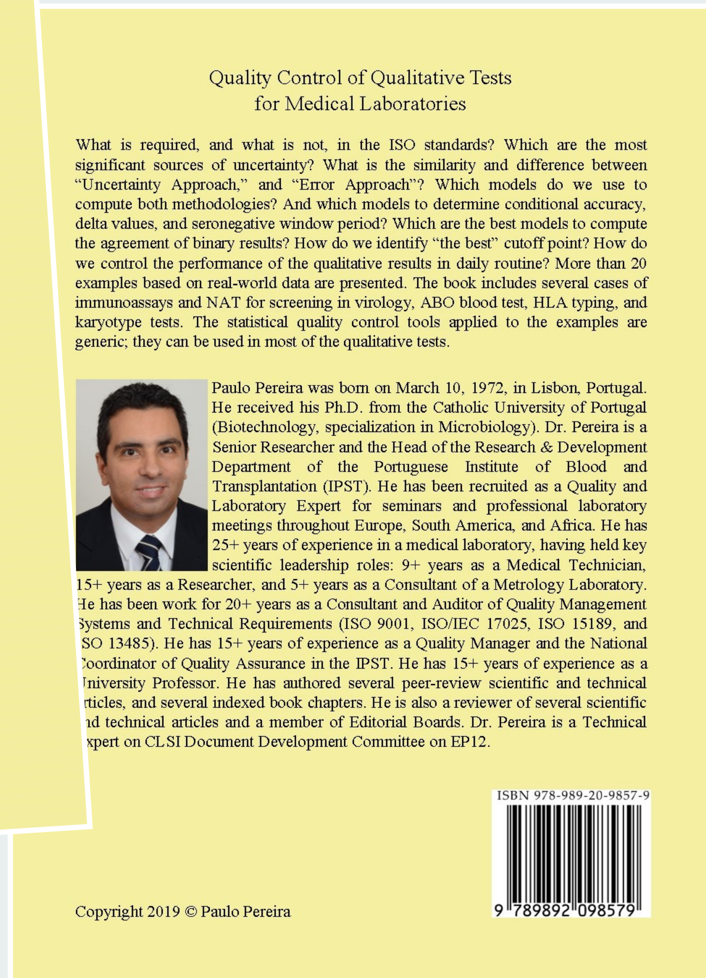
Quality Control of Qualitative Tests for Medical Laboratories

Introduction to the book

Paulo Pereira, Ph.D.

ISBN 978-989-20-9857-9





Quality Control of Qualitative Tests
for Medical Laboratories

What is required, and what is not, in the ISO standards? Which are the most significant sources of uncertainty? What is the similarity and difference between "Uncertainty Approach," and "Error Approach"? Which models do we use to compute both methodologies? And which models to determine conditional accuracy, delta values, and seronegative window period? Which are the best models to compute the agreement of binary results? How do we identify "the best" cutoff point? How do we control the performance of the qualitative results in daily routine? More than 20 examples based on real-world data are presented. The book includes several cases of immunoassays and NAT for screening in virology, ABO blood test, HLA typing, and karyotype tests. The statistical quality control tools applied to the examples are generic; they can be used in most of the qualitative tests.

Paulo Pereira was born on March 10, 1972, in Lisbon, Portugal. He received his Ph.D. from the Catholic University of Portugal (Biotechnology, specialization in Microbiology). Dr. Pereira is a Senior Researcher and the Head of the Research & Development Department of the Portuguese Institute of Blood and Transplantation (IPST). He has been recruited as a Quality and Laboratory Expert for seminars and professional laboratory meetings throughout Europe, South America, and Africa. He has 25+ years of experience in a medical laboratory, having held key scientific leadership roles: 9+ years as a Medical Technician, 15+ years as a Researcher, and 5+ years as a Consultant of a Metrology Laboratory. He has been work for 20+ years as a Consultant and Auditor of Quality Management Systems and Technical Requirements (ISO 9001, ISO/IEC 17025, ISO 15189, and ISO 13485). He has 15+ years of experience as a Quality Manager and the National Coordinator of Quality Assurance in the IPST. He has 15+ years of experience as a University Professor. He has authored several peer-review scientific and technical articles, and several indexed book chapters. He is also a reviewer of several scientific and technical articles and a member of Editorial Boards. Dr. Pereira is a Technical Expert on CLSI Document Development Committee on EP12.

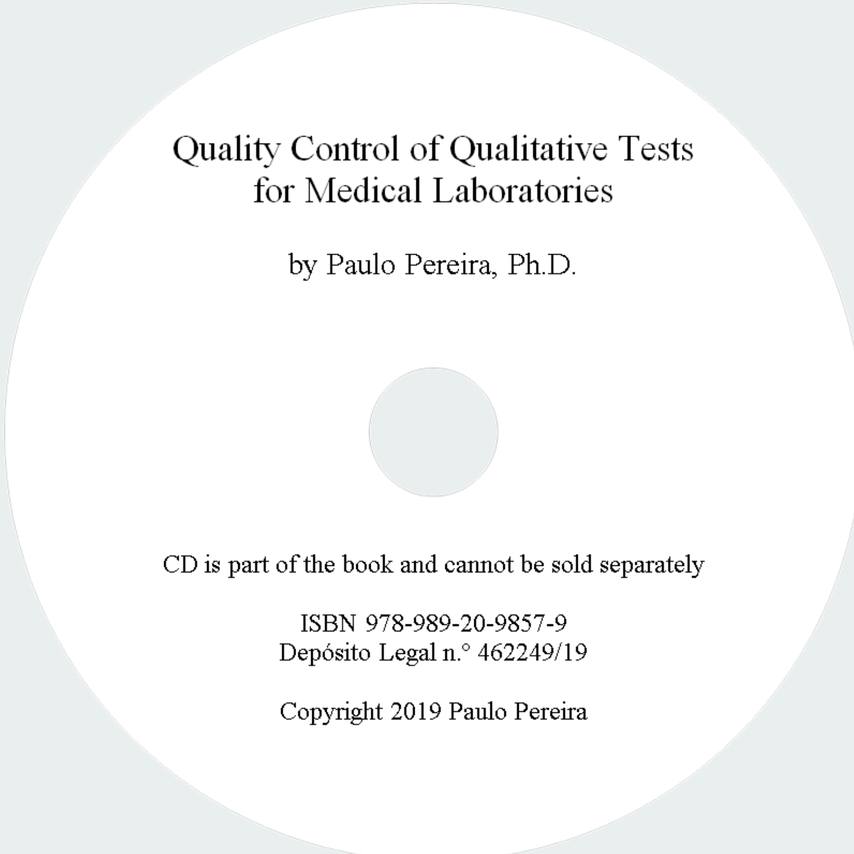
ISBN 978-989-20-9857-9

Copyright 2019 © Paulo Pereira

Why publish a book called "Quality control of qualitative tests for medical laboratories"?

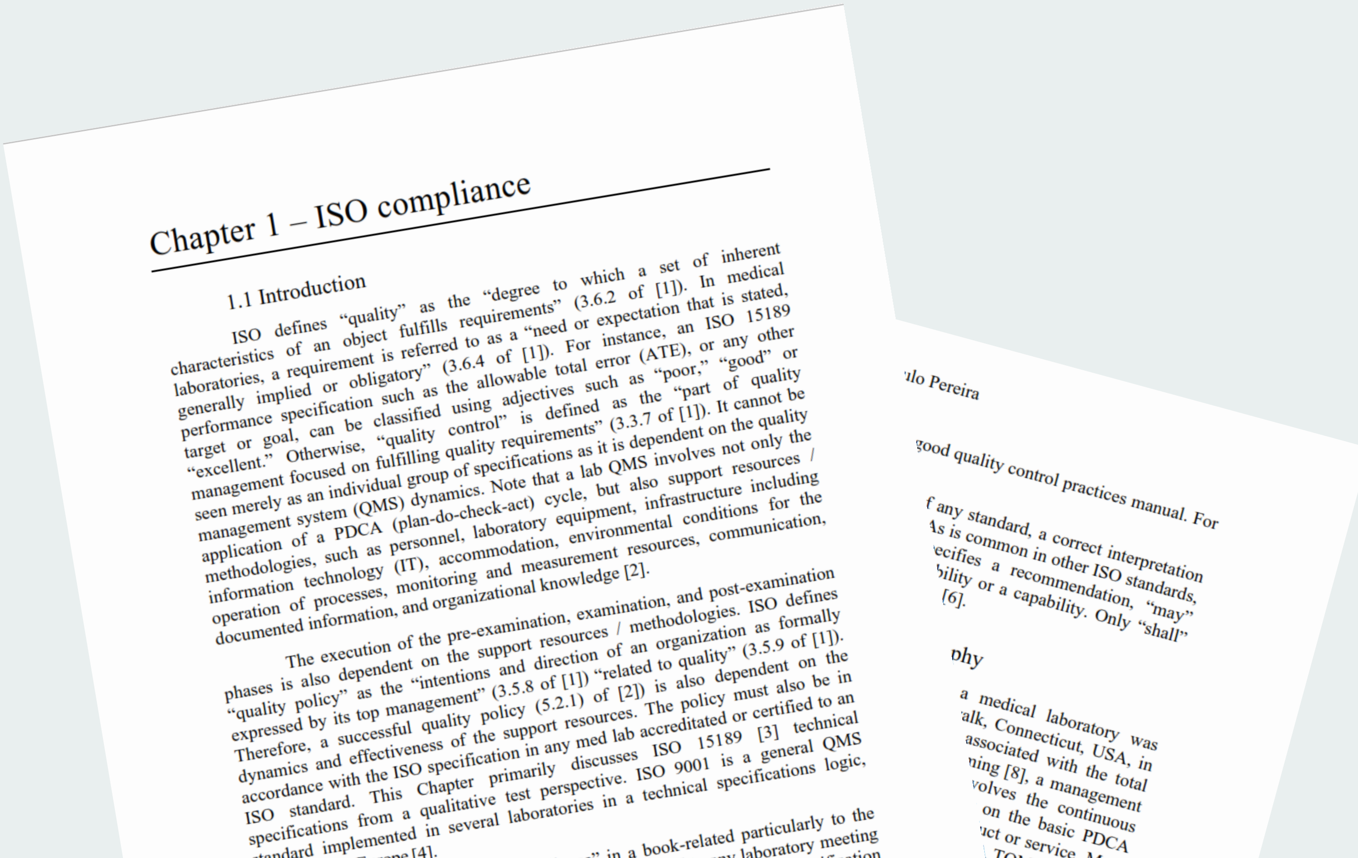
- Address the need for a book dedicated to quality control of qualitative tests
- The is a book written primarily for the laboratorian and aims to substantiate the selection of the best statistical tools considering the intended use of the qualitative tests' results (fitness for purpose)
- The purpose of the book is to answer most of qualitative tests QC questions in a three-pronged vision: the statistical, the clinical and the regulatory vision

- The book seeks to answer questions important to laboratory practice such as:
  - What is required, and what is not, in the ISO standards?
  - Which are the most significant sources of uncertainty?
  - What is the similarity and difference between “Uncertainty Approach,” and “Error Approach”?
  - Which models do we use to compute both methodologies?
  - And which models to determine conditional accuracy, delta values, and seronegative window period?
  - Which are the best models to compute the agreement of binary results?
  - How do we identify “the best” cutoff point?
  - How do we control the performance of the qualitative results in daily routine?
- More than 20 examples based on real-world data are presented
- The book includes several cases of immunoassays and NAT for screening in virology, ABO blood test, HLA typing, and karyotype tests
- The statistical quality control tools applied to the examples are generic; they can be used in most of the qualitative tests
- Approx. 200 pages printed on coated paper (couché) 90 grams; cover printed on 170 gram coated paper with soft-touch plastic coating; 2mm hard card cover



Why include a CD with spreadsheets?

- For a more natural comprehension of the approaches
- Facilitate the understanding of theory based on practice
- A practical way to demonstrate the case studies included in the book
- The laboratorian can easily replicate the models for his practice
- All the computations can be done using a conventional computer spreadsheet
- Excel® (Microsoft®, Redmond, Washington, USA) is immediately recognized as very intuitive software for laboratorian
- Readers will receive free updates to the spreadsheet package



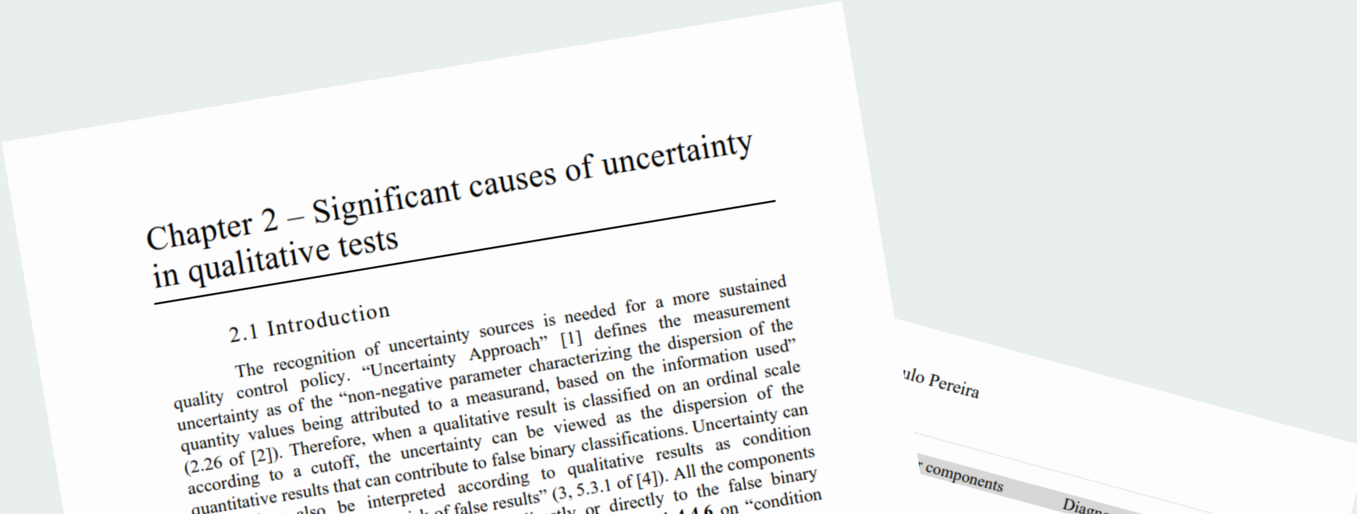
Chapter 1 – ISO compliance

1.1 Introduction

ISO defines "quality" as the "degree to which a set of inherent characteristics of an object fulfills requirements" (3.6.2 of [1]). In medical laboratories, a requirement is referred to as a "need or expectation that is stated, generally implied or obligatory" (3.6.4 of [1]). For instance, an ISO 15189 performance specification such as the allowable total error (ATE), or any other target or goal, can be classified using adjectives such as "poor," "good" or "excellent." Otherwise, "quality control" is defined as the "part of quality management focused on fulfilling quality requirements" (3.3.7 of [1]). It cannot be seen merely as an individual group of specifications as it is dependent on the quality management system (QMS) dynamics. Note that a lab QMS involves not only the application of a PDCA (plan-do-check-act) cycle, but also support resources / methodologies, such as personnel, laboratory equipment, infrastructure including information technology (IT), accommodation, environmental conditions for the operation of processes, monitoring and measurement resources, communication, documented information, and organizational knowledge [2].

The execution of the pre-examination, examination, and post-examination phases is also dependent on the support resources / methodologies. ISO defines "quality policy" as the "intentions and direction of an organization as formally expressed by its top management" (3.5.8 of [1]) "related to quality" (3.5.9 of [1]). Therefore, a successful quality policy (5.2.1) of [2]) is also dependent on the dynamics and effectiveness of the support resources. The policy must also be in accordance with the ISO specification in any med lab accreditated or certified to an ISO standard. This Chapter primarily discusses ISO 15189 [3] technical specifications from a qualitative test perspective. ISO 9001 is a general QMS standard implemented in several laboratories in a technical specifications logic, principally in Europe [4].

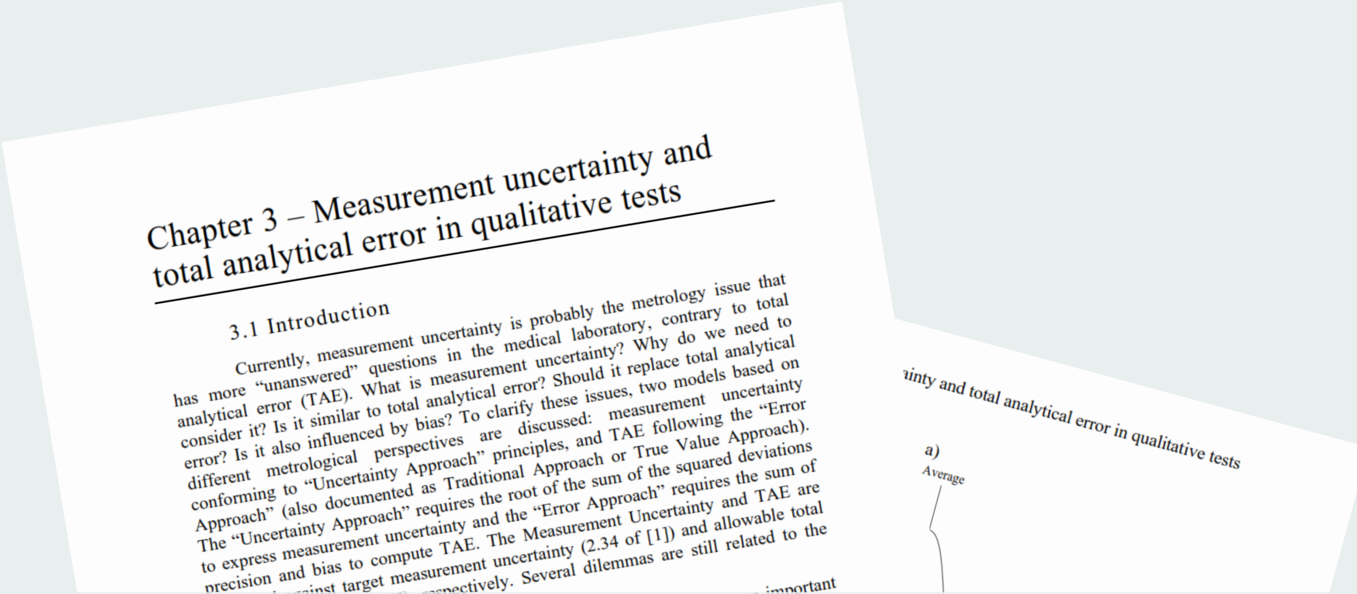
- **Chapter 1 – ISO compliance** introduces mainly ISO 15189 for the accreditation of medical laboratory methods or tests
- For a consistent application of this global standard, the laboratorian must understand its specifications
- We have discussed the use of most of its technical requirements that involve the selection, verification, validation, measurement uncertainty, internal quality control, and external quality assessment / proficiency testing (EQA / PT) of qualitative results
- Moreover, we have crossed ISO 15189 with ISO 9001 requirements for a more natural interpretation of this guideline, which is oriented to a generic implementation of a quality management system
- How do we meet the referred ISO claims? See the following chapters for suggested methodologies


Chapter 2 – Significant causes of uncertainty in qualitative tests

2.1 Introduction

The recognition of uncertainty sources is needed for a more sustained quality control policy. “Uncertainty Approach” [1] defines the measurement uncertainty as of the “non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used” (2.26 of [2]). Therefore, when a qualitative result is classified on an ordinal scale according to a cutoff, the uncertainty can be viewed as the dispersion of the quantitative results that can contribute to false binary classifications. Uncertainty can also be interpreted according to qualitative results as condition

- **Chapter 2 – Significant causes of uncertainty in qualitative tests** discusses the main sources of error that can cause untrue binary results
- As the test methodology is essential to recognize the most common analytical causes of failure, we have presented a brief overview of qualitative test design
- The impact of the analytical error on the cutoff trueness is discussed, as well as the effect of the analytical error on the accuracy of the classification of binary results
- The importance of the “gray zone” and the associated trinary classification to minimize the impact of analytical error in the results is debated
- The biased results due to biological factors are presented with a focus on the seroconversion window period
- The contribution of other possible sources of bias to the lack of representativeness of patients’ samples is also pondered.
- The impact of interferences in bias is discussed
- This debate is important for a better focus on the use of the quality control tools that allow us to see what is and what is not measurable (limitation of the studies)

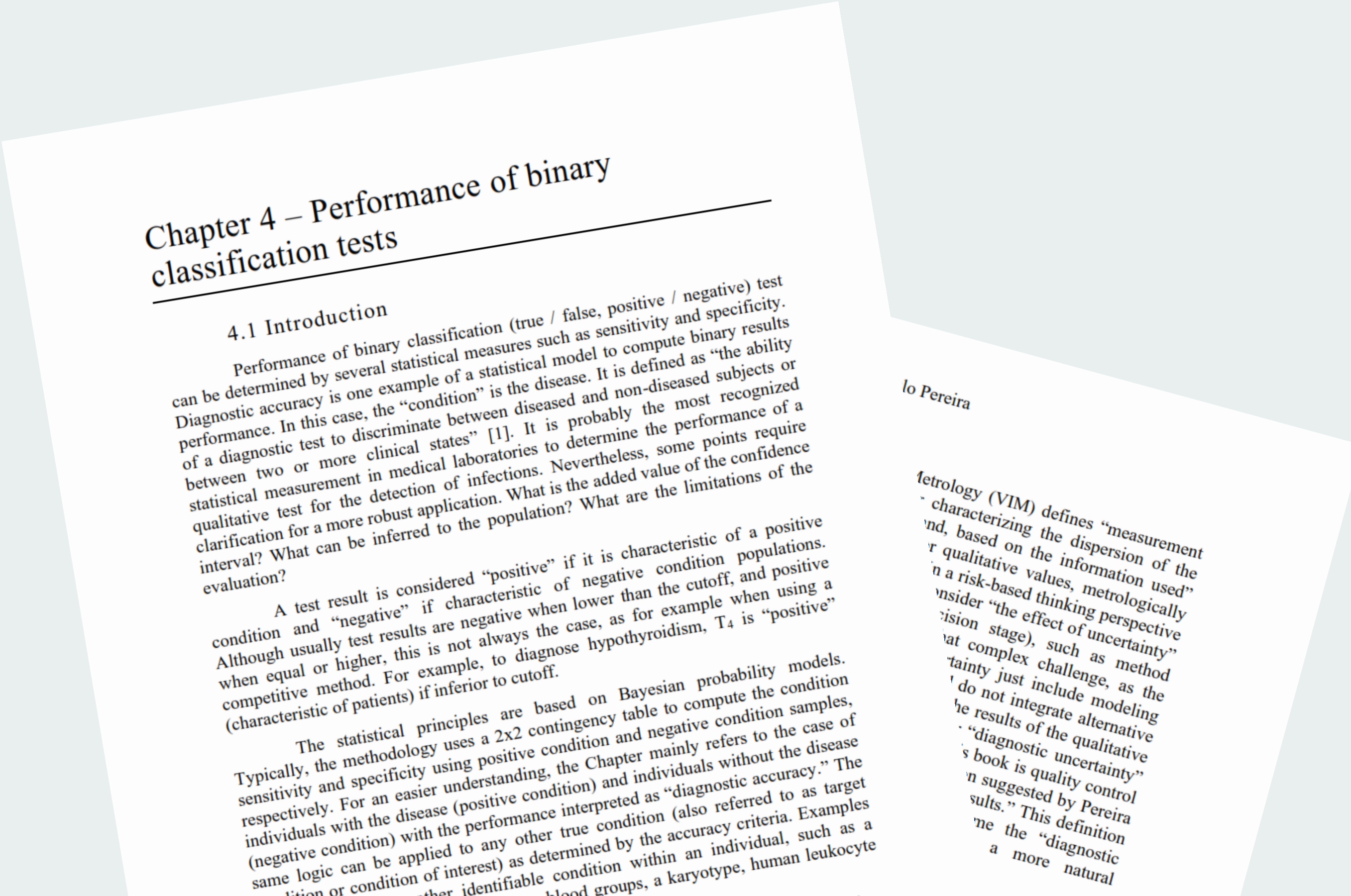


Chapter 3 – Measurement uncertainty and total analytical error in qualitative tests

3.1 Introduction

Currently, measurement uncertainty is probably the metrology issue that has more "unanswered" questions in the medical laboratory, contrary to total analytical error (TAE). What is measurement uncertainty? Why do we need to consider it? Is it similar to total analytical error? Should it replace total analytical error? Is it also influenced by bias? To clarify these issues, two models based on different metrological perspectives are discussed: measurement uncertainty conforming to "Uncertainty Approach" principles, and TAE following the "Error Approach" (also documented as Traditional Approach or True Value Approach). The "Uncertainty Approach" requires the root of the sum of the squared deviations to express measurement uncertainty and the "Error Approach" requires the sum of precision and bias to compute TAE. The Measurement Uncertainty and TAE are evaluated against target measurement uncertainty (2.34 of [1]) and allowable total error (1.5.2 of [2]), respectively. Several dilemmas are still related to the

- **Chapter 3 – Measurement uncertainty and total analytical error in qualitative methods** introduces both the Uncertainty Approach" and the "Error Approach"
- The challenge is to introduce the laboratorian to the similarities and differences of the visions, wherein empirical models are considered for both visions
- While not ignoring the usefulness of the modular models to the manufacturer, they are not discussed further here since they are not meant to be used in medical laboratory practice
- The models presented are based on recognized protocols in med lab requiring data from single-laboratory validation, interlaboratory comparisons or EQA /PT
- The importance of the metrological traceability of the results is considered
- Compliance assessment is associated with the empirical estimate of the "gray zone" and the limit of detection (LoD)
- The evaluation of analyte concentrations near the cutoff is presented as a complementary tool to estimate an identical zone



Chapter 4 – Performance of binary classification tests

4.1 Introduction

Performance of binary classification (true / false, positive / negative) test can be determined by several statistical measures such as sensitivity and specificity. Diagnostic accuracy is one example of a statistical model to compute binary results performance. In this case, the "condition" is the disease. It is defined as "the ability of a diagnostic test to discriminate between diseased and non-diseased subjects or between two or more clinical states" [1]. It is probably the most recognized statistical measurement in medical laboratories to determine the performance of a qualitative test for the detection of infections. Nevertheless, some points require clarification for a more robust application. What is the added value of the confidence interval? What can be inferred to the population? What are the limitations of the evaluation?

A test result is considered "positive" if it is characteristic of a positive condition and "negative" if characteristic of negative condition populations. Although usually test results are negative when lower than the cutoff, and positive when equal or higher, this is not always the case, as for example when using a competitive method. For example, to diagnose hypothyroidism, T_4 is "positive" (characteristic of patients) if inferior to cutoff.

The statistical principles are based on Bayesian probability models. Typically, the methodology uses a 2x2 contingency table to compute the condition sensitivity and specificity using positive condition and negative condition samples, respectively. For an easier understanding, the Chapter mainly refers to the case of individuals with the disease (positive condition) and individuals without the disease (negative condition) with the performance interpreted as "diagnostic accuracy." The same logic can be applied to any other true condition (also referred to as target condition or condition of interest) as determined by the accuracy criteria. Examples of condition or any other identifiable condition within an individual, such as a ...

lo Pereira

...etrology (VIM) defines "measurement ... characterizing the dispersion of the ...nd, based on the information used" ...r qualitative values, metrologically ...n a risk-based thinking perspective ...nsider "the effect of uncertainty" ...cision stage), such as method ...at complex challenge, as the ...tainty just include modeling ... do not integrate alternative ...he results of the qualitative ... "diagnostic uncertainty" ...s book is quality control ...n suggested by Pereira ...sults." This definition ...ne the "diagnostic ... a more natural

- **Chapter 4 – Performance of binary classification tests** is based on condition accuracy, probably the most well-known methodology for validating qualitative results
- In this chapter, we introduce the basis of the statistics concepts applied and discuss the importance of the samples to the robustness of the estimates
- We have used 2x2 contingency tables, followed by a discussion about the value of the analysis of the numerical data to distinguish between two or more tests with identical condition sensitivity and specificity
- The concept of "condition uncertainty" is introduced, analogous to the "measurement uncertainty" of quantitative dimensions
- The window period is presented using a binary and trinary

Chapter 5 – Agreement of binary classification tests

5.1 Introduction

Every so often, samples with the known condition for evaluating qualitative tests are not available, especially those intended to detect rare conditions, such as diseases. As an alternative, samples with known results in a comparator test are suggested (10.2 of [1]). A disadvantage is that it is not possible to compute condition accuracy rates (see **Chapter 4**). Sometimes the agreement is misunderstood with condition accuracy such as diagnostic accuracy. In this case, a bad practice should be immediately recognized, as condition accuracy rates require evidenced-based condition information as diagnostic info. Condition sensitivity and specificity cannot be determined just by the results of a test that is not a "gold standard" (see **4.3.1**). Misinterpretation of the accuracy of results of a new test could, therefore, happen, and the correct use of the binary may be compromised. If the positive condition and negative condition samples are not available, one could ask a few questions. How do we determine the agreement of binary results? Does the confidence interval add value to the evaluation? Can the agreement be inferred to the population? What are the limitations of the study?

The Chapter presents a set of good practices to guarantee reliable agreement evaluations. These practices are suited to support the robust use of agreement methods. Real-world data is used to illustrate the models.

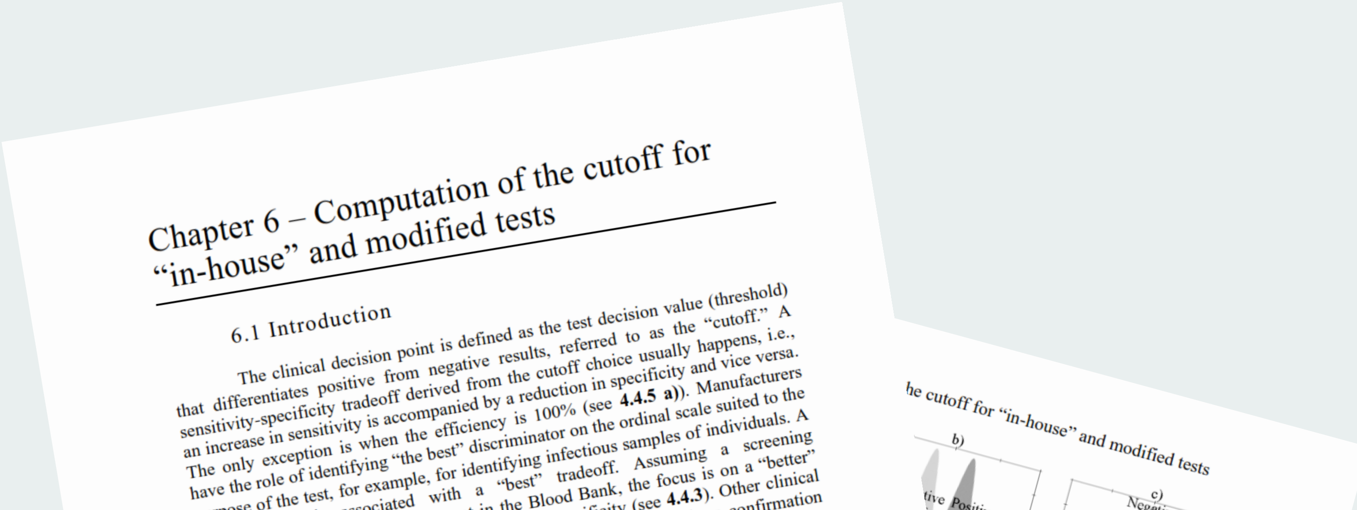
5.2 Principles

The purpose of the agreement of binary results is to evaluate a test based on the agreement of the results of a candidate and a comparative test. The comparator results are assumed as those with the highest probability of being true. Thus, the accuracy of the study depends on the comparative test performance. This test should not be confused with a "gold standard." Thus, the comparator should show at least state-of-the-art performance, i.e., a demonstrated top performance of condition accuracy. This information can be collected in a review of the literature for most of the assays. The candidate test performance decreases proportionally with the number of discrepant results against the comparative assay. Then, the selection of the comparator is critical to the reliability and consistency of the evaluation. For example, the U.S. Food and Drug Administration (FDA) ranking (V.C.1 of [2]) classifies the quality of a comparative method as follows:

- type A as "a quantitative reference method (…) with the appropriate cutoff value for the positive and negative results";
- type B as "a qualitative reference method (…)";

111

- **Chapter 5 – Agreement of binary classification tests** is intended to lead the reader to validation where samples with a true condition are unavailable
- Since the consistency of the results is dependent on the comparative test performance, its selection should be applied uniquely if the condition is unknown



Chapter 6 – Computation of the cutoff for “in-house” and modified tests

6.1 Introduction

The clinical decision point is defined as the test decision value (threshold) that differentiates positive from negative results, referred to as the “cutoff.” A sensitivity-specificity tradeoff derived from the cutoff choice usually happens, i.e., an increase in sensitivity is accompanied by a reduction in specificity and vice versa. The only exception is when the efficiency is 100% (see **4.4.5 a)**). Manufacturers have the role of identifying “the best” discriminator on the ordinal scale suited to the purpose of the test, for example, for identifying infectious samples of individuals. A “better” cutoff is associated with a “best” tradeoff. Assuming a screening

- **Chapter 6 – Computation of the cutoff for “in-house” and modified tests**, as the title refers, applies solely to tests prepared in the laboratory requiring cutoff determination
- The “realism” of the cutoff does not depend only on the samples but also on the intended use of the results
- Usually, false-positive results are better accepted than false-negative ones
- The computation of the cutoff by the receiver operating characteristic curve (ROC) is discussed
- Although we have tried to use the most accessible language, it is probably the most complex statistical model presented in this book
- However, its principle is simple: it provides the various condition sensitivities and specificities for all the possible cutoff points
- The laboratorian selects the point that meets the requirements related to the intended use of the results, i.e., according to the clinical application
- An area ranking allows the classification of the detection capability of the test for a certain cutoff

- **Chapter 7 – Internal quality control and external quality assessment / proficiency testing** debate models suitable for qualitative tests
- The internal quality control principles are discussed to aid the selection of the best designs based on a qualitative logic
- Demystification of control rules in qualitative tests statistically and clinically supported
- Novel approaches to compute sigma metrics in qualitative tests
- The DPMO-derived and SE_{crit}-derived sigma metrics express the capability of tests to meet the specifications
- Models are presented for variables using numerical results (ordinal tests), and an application to monitor “pure” qualitative results (nominal tests)
- Both methodologies are intended to control the loss of sensitivity in the qualitative tests
- EQA /PT is introduced