Clinica Chimica Acta 430 (2014) 1–8

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

Influence of analytical bias and imprecision on the number of false positive results using Guideline-Driven Medical Decision Limits

Per Hyltoft Petersen [a,*], George G. Klee [b]

[a] The Norwegian Quality Improvement of Primary Care Laboratories (NOKLUS), Section for General Practice, Department of Public Health and Primary Health Care, Faculty of Medicine and Dentistry, University of Bergen, Norway
[b] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

ARTICLE INFO

Article history:
Received 13 June 2013
Received in revised form 23 October 2013
Accepted 11 December 2013
Available online 18 December 2013

Keywords:
Analytical bias
Analytical imprecision
Analytical quality specifications
Clinical guidelines
Clinical outcome
Probability function

ABSTRACT

Background: Diagnostic decisions based on decision limits according to medical guidelines are different from the majority of clinical decisions due to the strict dichotomization of patients into diseased and non-diseased. Consequently, the influence of analytical performance is more critical than for other diagnostic decisions where much other information is included. The aim of this opinion paper is to investigate consequences of analytical quality and other circumstances for the outcome of "Guideline-Driven Medical Decision Limits".
Terms: Effects of analytical bias and imprecision should be investigated separately and analytical quality specifications should be estimated accordingly.
Biological variation and analytical performance: Use of sharp decision limits doesn't consider biological variation and effects of this variation are closely connected with the effects of analytical performance. Such relationships are investigated for the guidelines for HbA_{1c} in diagnosis of diabetes and in risk of coronary heart disease based on serum cholesterol. The effects of a second sampling in diagnosis give dramatic reduction in the effects of analytical quality showing minimal influence of imprecision up to 3 to 5% for two independent samplings, whereas the reduction in bias is more moderate and a 2% increase in concentration doubles the percentage of false positive diagnoses, both for HbA_{1c} and cholesterol.
Frequency of follow-up laboratory tests: An alternative approach comes from the current application of guidelines for follow-up laboratory tests according to clinical procedure orders, e.g. frequency of parathyroid hormone requests as a function of serum calcium concentrations. Here, the specifications for bias can be evaluated from the functional increase in requests for increasing serum calcium concentrations.
Probability function for diagnoses: In consequence of the difficulties with biological variation and the practical utilization of concentration dependence of frequency of follow-up laboratory tests already in use, a kind of probability function for diagnosis as function of the key-analyte is proposed.

1. Introduction

In 1999 an international consensus conference on "Strategies to Set Global Analytical Quality Specifications in Laboratory Medicine" was held in Stockholm [1], where the clinical chemical scientists within the field of analytical goal-setting agreed on the following hierarchy of acceptable models which should be used to set analytical quality specifications [2]:

1. Evaluation of the effect of analytical performance on clinical outcomes in specific clinical situations
2. Evaluation of the effect of analytical performance on clinical decisions in general
   a. Data based on the components of biological variation
   b. Data based on analysis of clinicians' opinions
3. Published professional recommendations
   a. From national and international expert bodies
   b. From expert local groups or individuals
4. Performance goals set by
   a. Regulatory bodies
   b. Organisers of External Quality Assessment (EQA) schemes
5. Goals based on the current state of the art
   a. As demonstrated by data from EQA or Proficiency Testing Schemes
   b. As found in current publications on methodology

Before this conference, there were competing approaches to goal-setting in clinical chemistry, which made it difficult to decide the

* Corresponding author at: Norwegian quality improvement of primary care laboratories, Division for General Practice, University of Bergen Box 6165, 5892 Bergen, Norway; Flittig Lise Vej 20, DK-5250 Odense SV, Denmark. Tel.: +45 65 96 25 65.
E-mail address: Per.Petersen@isf.uib.no (P. Hyltoft Petersen).

0009-8981/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cca.2013.12.014

relevant strategy for estimation of analytical quality specifications in specific projects and clinical situations.

This consensus agreement made it easier to decide strategy and analytical quality specifications in clinical chemistry as also demonstrated in the follow-up ten years later [3] but here, it was also concluded that analytical quality specifications for areas such as matrix effects and measurements on ordinal scale still need further investigation to produce relevant analytical quality specifications.

In 2010 George Klee [4] gave a proposal for modifying and expanding the hierarchical system to six approaches for establishing outcome-related performance goals:

a) limits defined by regulations and external assessment programmes
b) limits based on biological variation
c) limits based on surveys of clinicians about their needs
d) limits based on their effects on guideline-driven medical decisions
e) limits based on analysis of patterns for ordering follow-up clinical tests
f) limits based on formal medical decision models

In this review the order is somewhat changed compared with the consensus. Section a) in the review corresponds to the consensus number 4. Section b) corresponds to consensus number 2a and section c) corresponds to consensus 2b. Sections d), e), and f) relate to consensus number 1; however, the review has used surrogate measures for "clinical outcomes in specific clinical situations", since very little data are available which quantify the effects of analytic error on clinical outcomes. This review evaluates the effects of analytic error on guideline-driven medical decisions, the ordering of follow-up clinical tests, and formal medical decision models.

Section e) 'analysis of patterns for ordering follow-up clinical tests' is illustrated by an analysis of the effect of serum calcium results indicating hypercalcemia on follow-up requests for serum PTH measurements for evaluating possible diseases of the parathyroid. Section f) 'analytical performance characteristics based on decision models used in expert systems' is illustrated by a 'cost' model of decision function for TSH concentrations to classify patients into hypothyroid, normal thyroid and hyperthyroid states.

Section d) 'analytical performance characteristics based on their effects on guideline-driven medical decisions', is explained by the guidelines on increased risk for coronary artery disease using serum cholesterol measurements. This section also introduces a new concept that decision limits may be variable [5] Fig. 1. Many guidelines include very specific decision limits, such as defined concentrations of key laboratory tests. In practice, however, the rationale of the guideline may be followed, but the specific decision level may be adjusted to account for individual patient circumstances or individual medical centre preferences. This variation in decision limits would be a logical mechanism to adjust for assay method and calibration differences, although these differences seldom are explicitly acknowledged in most guideline applications. Fig. 1 shows a series of individual decisions as dotted step functions, with the probability function going from zero to 100% at specific decision levels. The solid sigmoid line represents a composite integration of multiple decisions. The tracking of ordering patterns of actual clinical decisions for follow-up procedures as a function of the levels of key analytic assays shows sigmoid patterns, similar to the one illustrated in the figure.

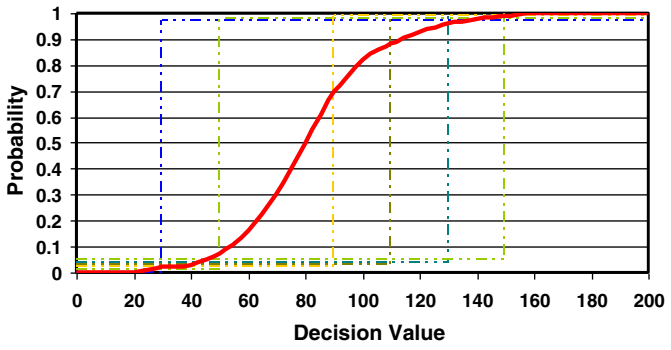

Fig. 1. Comparison of the step function probability functions of multiple patient specific decision limits (shown as dashed lines) versus the sigmoidal decision function (solid curve) representing the integrated composite of the decisions for all the patients.

The same statistical analyses presented for single decision points in this current manuscript, can be generalized to apply to decisions with varying decision limits and for other analytical components as well.

The purpose of this contribution is to perform a detailed analysis of the influence of one or two samplings and analytical performance on two guideline-driven medical decisions. One for the diagnosis of diabetes using Haemoglobin A1c, HbA_{1c}, and one for the classification of patients with low-risk and high-risk for coronary heart disease based on serum cholesterol. Further, we want to discuss the relevance of sharp decision limits to be used in diagnosis of individuals and to compare this with the overall interpretation of clinical outcome from large population studies as reflected in the clinical guidelines. This concept is illustrated with clinical data showing that the relative frequency of laboratory orders for parathyroid testing does not increase as a step function at the guideline specified value, but increases progressively as the concentration of calcium becomes more abnormal.

2. Basic concepts for allowable analytical bias and imprecision, and possible combinations

2.1. Terms

In the routine clinical chemistry laboratory we are familiar with the meaning of analytical bias and analytical imprecision even though these concepts are hardly defined in VIM (International Vocabulary of Metrology—basic and general concepts and associated terms) [6] where bias shortly is defined as "estimate of a systematic measurement error" and imprecision is a note under "measurement precision" which is "usually expressed numerically by measures of imprecision, such as….coefficient of variation". We will keep the short laboratory terms in the following and only distinguish between permanent (method) bias as well as the inherent within-run and total (laboratory) imprecision. Further, we only describe analytical measurement results performed on a ratio scale, which makes it possible to measure/estimate fractional or percentage differences and to calculate coefficients of variation (CV-values) as fractions or as percentages.

2.2. Models for combination of bias and imprecision

Systematic errors and random errors are by definition different, but varying general models have been proposed for the combination of the two types of errors in relation to analytical quality specifications [7].

2.2.1. The variance model

This model is the traditional variance model, where both bias and imprecision are treated as variables in the formula: $\sigma_{Combined}^2 = \sigma_{Laboratory}^2 + |bias|^2$, where bias is squared [8]. In this model the maximum bias is equal to the maximum imprecision: $\sigma_{Max} = |bias_{Max}|$ when the other is zero.

2.2.2. The GUM model

This model is based on the theory on estimation of uncertainty of measurements [9], where an unknown systematic error is assumed to be rectangularly distributed with same probability for any value of the presumed interval, 2^*A. This rectangular distribution is then transformed to a standard deviation $\sigma_{Interval} = A/3^{½}$ or $A = 3^{½*}\sigma_{Interval} = 1.73^*\sigma_{Interval}$, where A in reality is the same as