Original Investigation

Interpreting Change in Quantitative Imaging Biomarkers

Nancy A. Obuchowski, PhD

Rationale and Objectives: Quantitative imaging biomarkers (QIBs) are becoming increasingly adopted into clinical practice to monitor changes in patients' conditions. The repeatability coefficient (RC) is the clinical cut-point used to discern between changes in a biomarker's measurements due to measurement error and changes that exceed measurement error, thus indicating real change in the patient. Imaging biomarkers have characteristics that make them difficult for estimating the repeatability coefficient, including nonconstant error, non-Gaussian distributions, and measurement error that must be estimated from small studies.

Methods: We conducted a Monte Carlo simulation study to investigate how well three statistical methods for estimating the repeatability coefficient perform under five settings common for QIBs.

Results: When the measurement error is constant and replicates are normally distributed, all of the statistical methods perform well. When the measurement error is proportional to the true value, approaches that use the log transformation or coefficient of variation perform similarly. For other common settings, none of the methods for estimating the repeatability coefficient perform adequately.

Conclusion: Many of the common approaches to estimating the repeatability coefficient perform well for only limited scenarios. The optimal approach depends strongly on the pattern of the within-subject variability; thus, a precision profile is critical in evaluating the technical performance of QIBs. Asymmetric bounds for detecting regression vs progression can be implemented and should be used when clinically appropriate.

Key Words: Quantitative imaging biomarker; repeatability; repeatability coefficient; measurement error.

© 2017 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

INTRODUCTION

As quantitative imaging biomarkers (QIBs) become increasingly adopted into clinical practice for diagnosis, prognosis, and disease monitoring, it is critical that clinicians be able to interpret them properly. Clinicians must be able to discern between changes in a biomarker's measurements that are expected because of measurement error by the imaging system and changes that exceed measurement error and thus indicate a real change in the patient.

Studies of the technical performance of QIBs, particularly studies of their repeatability, are used to help radiologists interpret change. Organizations such as the Quantitative Imaging Biomarker Alliance (QIBA) conduct groundwork studies to estimate biomarkers' performance. They also perform meta-analyses to summarize biomarkers' performance over multiple published studies (1,2). From these studies of repeatability, investigators calculate a cut-point, or threshold, for discerning when a measured change is attributable to measurement error vs when the measured change should be interpreted as a real change in the patient, with some stated degree of confidence (3–5). For illustration, when measuring the change in volume of a pulmonary lesion from baseline, a measured change <25% might be considered purely measurement error, whereas a measured change exceeding 25% might be considered a real change, with 95% confidence (6).

In this paper, we present the results of a Monte Carlo simulation study investigating how well these clinical cut-points perform with respect to their declared confidence levels. We consider scenarios common for imaging biomarkers, including nonconstant measurement error, non-Gaussian distributions, and small technical performance studies from which the cut-points are estimated. We consider three statistical methods for estimating the cut-point. We investigate the performance of the cut-points when no real change has occurred and also when a real change has occurred (both regression and progression). Based on these findings, we provide recommendations for how best to estimate the cut-point for interpreting change in QIBs.

METHODS

Defining Change

There are three common ways in which change in a quantitative biomarker can be defined. Let Y_{ib} be the biomarker measurement taken on the ith subject at baseline and let Y_{it} be the biomarker measurement taken on the same subject at some follow-up timepoint. Table 1 summarizes three ways to define change: (1) difference from baseline expressed in

Acad Radiol 2017; ■:■■–■■

From the Quantitative Health Sciences/JJN3, Cleveland Clinic Foundation, 9500 Euclid Ave, Cleveland, OH 44195-5196. Received August 30, 2017; revised September 26, 2017; accepted September 27, 2017. **Address correspondence to:** N.A.O. e-mail: obuchon@ccf.org

© 2017 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.acra.2017.09.023

TABLE 1. Three Definitions of Change in QIBs*

Difference from baseline	$\hat{d} = (Y_{it} - Y_{ib})$	(1)
Percentage change from baseline	$\%\hat{d}_b = (Y_{it} - Y_{ib})/Y_{ib} \times 100$	(2)
Percentage difference	$\%\hat{d}_{mean} = \frac{(Y_{it} - Y_{ib})}{(Y_{ib} + Y_{it})/2} \times 100$	(3)

QIBs, quantitative imaging biomarkers.

* Positive values are indicative of an increase in the biomarker value at follow-up (referred to here as "progression" without loss of generality) and negative values are indicative of a decrease in the biomarker value at follow-up ("regression").

the original units of the biomarker, $\hat{d}$; (2) percentage change from baseline, $\%\hat{d}_b$; and (3) percentage difference in the two measurements, $\%\hat{d}_{mean}$. We will examine cut-points for each of these in our study.

Methods for Estimating Cut-Points From Technical Performance Studies

The biomarker measurements are assumed to be a function of the true biomarker value, with some possible bias and measurement error. Let X_{ik} be the true biomarker value at timepoint k for subject i. The biomarker measurement is written as a linear function of the true value (3):

$$Y_{ik} = \beta_o + \beta_1 X_{ik} + \epsilon_{ik}, \quad (4)$$

where β_o is the fixed bias, β_1 is the proportional bias, and ϵ_{ik} is the measurement error. It is often assumed that the measurement error is constant and follows a normal distribution, $\epsilon_{ik} \sim N(0, \sigma^2)$. We vary these assumptions in our Monte Carlo simulation study to consider scenarios typical in imaging. The measurement error, σ^2, is often referred to as the within-subject variance. Its square root is referred to as the within-subject standard deviation (wSD) (3,4).

To estimate the cut-point at which changes in the biomarker measurements can be considered real changes, technical performance studies are conducted to estimate σ^2, the within-subject variance. These studies, often called test-retest studies, are designed so that a subject is imaged in the same fashion on multiple occasions over a short period of time to ensure that no biological change has occurred (4). As it is often difficult to justify imaging a subject multiple times over a short period of time (often the same day), these studies are usually limited to two replicates per subject (3,4). Obuchowski and Bullen (7) recommended that at least 35 test-retest subjects are needed for providing 95% coverage of the true biomarker value of future subjects. For a test-retest study with N subjects and two replicates per subject, an estimate of the within-subject variance is

$$\hat{\sigma}^2 = \sum_{i=1}^{N} \{Y_{i1} - Y_{i2}\}^2 / 2N, \quad (5)$$

where Y_{ik} denotes the biomarker measurement for the kth replicate on the ith subject. Assuming constant within-subject variance and normality, the cut-point associated with 95% confidence is then

$$\widehat{RC} = 1.96 \times \sqrt{2\hat{\sigma}^2}, \quad (6)$$

where RC is the repeatability coefficient, or the least significant difference between two repeat measurements taken under identical conditions (3–5,8). Changes in the biomarker that exceed +RC or are less than −RC are considered real changes (not as a result of just measurement error), with 95% confidence. Specifically, measured differences <−RC are considered true regression, whereas measured differences >+RC are considered true progression. We refer to this simple approach as Approach 1 (see first row of Table 2).

The cut-points in Approach 1 are symmetric, such that a similar magnitude of progression or regression is considered true change (Fig 1). Consider an example for lung tumor volume measurements. From a previous technical performance study, the RC was estimated as 150 mm^3 (ie, $\hat{\sigma} = 54$). Now, suppose that in a future subject the measured volume of a lung nodule is 500 mm^3 at baseline, and after treatment, the measured volume is 300 mm^3. The measured change is (300–500), or a 200 mm^3 reduction in volume. Is this difference due to measurement error or should we conclude that a real change has occurred? Because $-200 < -\widehat{RC}$, we would conclude that a real change has occurred with 95% confidence.

It is common for biomarkers' wSD to be correlated with the magnitude of the biomarker (8). When there is a relationship between the wSD and the magnitude of the measurements, there are two common approaches to minimizing this relationship (2,3,8): (1) within-subject coefficient of variation (wCV) approach and (2) log transformation. These approaches lead to symmetric and asymmetric intervals, respectively (Fig 1). We discuss these approaches next.

From the technical performance study, the wCV is estimated by first calculating the mean of the replicate measurements for each of the N subjects. Denote the mean for the ith subject as $\bar{Y}_i = (Y_{i1} + Y_{i2})/2$. The estimate of the wCV^2 is

$$\widehat{wCV}^2 = \sum_{i=1}^{N} \left\{ (Y_{i1} - Y_{i2})^2 / (2\bar{Y}_i^2) \right\} / N. \quad (7)$$

The $\%RC$ is estimated as

$$\%\widehat{RC} = 1.96 \times \sqrt{2\widehat{wCV}^2}$$

and the symmetric cut-points are defined as Approach 2 in Table 2. Note that $\%RC$ could be used to define a cut-point for the percentage change from baseline ($\%\hat{d}_b$) or for the percentage difference ($\%\hat{d}_{mean}$). These are presented as Approaches 2A and 2B in Table 2, respectively.

In the log transformation approach, the variance of the log-transformed data is calculated as (8,9)

$$\hat{\sigma}^2_{\ln} = \sum_{i=1}^{N} \{\ln(Y_{i1}) - \ln(Y_{i2})\}^2 / 2N. \quad (8)$$

Download English Version:

https://daneshyari.com/en/article/8821006

Download Persian Version:

https://daneshyari.com/article/8821006

Daneshyari.com