Published in final edited form as:

Stat Med. 2013 May 30; 32(12): 2048–2061. doi:10.1002/sim.5668.

Standard error of inverse prediction for dose–response relationship: approximate and exact statistical inference

Eugene Demidenko^{*,†}, Benjamin B. Williams, Ann Barry Flood, and Harold M. Swartz
Dartmouth Physically Based Biodosimetry Center for Medical Countermeasures Against Radiation (Dart-Dose CMCR), Geisel School of Medicine at Dartmouth, Hanover, NH 03755, U.S.A

Abstract

This paper develops a new metric, the standard error of inverse prediction (SEIP), for a dose–response relationship (calibration curve) when dose is estimated from response via inverse regression. SEIP can be viewed as a generalization of the coefficient of variation to regression problem when x is predicted using y -value. We employ nonstandard statistical methods to treat the inverse prediction, which has an infinite mean and variance due to the presence of a normally distributed variable in the denominator. We develop confidence intervals and hypothesis testing for SEIP on the basis of the normal approximation and using the exact statistical inference based on the noncentral t -distribution. We derive the power functions for both approaches and test them via statistical simulations. The theoretical SEIP, as the ratio of the regression standard error to the slope, is viewed as reciprocal of the signal-to-noise ratio, a popular measure of signal processing. The SEIP, as a figure of merit for inverse prediction, can be used for comparison of calibration curves with different dependent variables and slopes. We illustrate our theory with electron paramagnetic resonance tooth dosimetry for a rapid estimation of the radiation dose received in the event of nuclear terrorism.

Keywords

dose–response; calibration; coefficient of variation; confidence interval; coverage probability; nuclear terrorism; power function

1. Introduction

Dose–response relationship, or calibration curve, is an established quantitative tool of science and technology. Typically, response y is related to dose/stimulus x via simple linear regression

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad (1)$$

where y_i is the i th observation of the response to dose x_i in n experiments, $i = 1, 2, \dots, n$. Here α and β are unknown regression parameters (α is the intercept, and β is the slope), and ε_i is an unobservable error term with zero mean and constant variance σ^2 . In this paper, we assume that x_i are controlled/fixed dose values and $\{\varepsilon_i, i = 1, \dots, n\}$ are independent and

identically normally distributed as $\mathcal{N}(0, \sigma^2)$. Parameter σ is referred to as regression standard error.

In many studies, regression (1) is used to predict y given x , generally referred to as *direct prediction*. For example, in toxicology, one wants to predict the mortality of pest (y) given pesticide concentration (x). In this case, regression standard error σ serves as the metric of the accuracy prediction $\hat{y} = a + bx$ given x because

$$SE(\hat{y}) = \sigma \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (2)$$

where a and b are the least squares estimates for α and β , respectively. Consequently, in direct prediction problems, we are looking for the best fit that is equivalent to minimizing the residual sum of squares, so that the respective metric that reflects the quality of the dose–response relationship, the standard error of prediction (SEP), is σ .

In other studies, the inverse of this problem is of interest, referred to as *inverse prediction*; that is, the response y is known, and the variable to be predicted is x [1, p. 172]. The example used in this paper (see more detail in Section 5) is that we measure the reaction of tooth enamel to radiation and want to estimate the dose received from accidental or terrorist-related exposures. Thus, in inverse prediction, we predict dose x given response value y as

$$\hat{x} = \frac{y - a}{b}, \quad (3)$$

sometimes called as the method of inverse regression or classical calibration problem [2]. An alternative approach, when x is predicted from regression of x on y , is called inverse estimator approach [3]. The discussion of the pros and cons of the two approaches is deferred to the end of the paper.

Statistical inference for prediction (3) is not trivial because of the presence of a normally distributed random variable (b) in the denominator. Because $\hat{x}$ is the ratio of two normally distributed random variables, its distribution follows the Cauchy distribution and consequently has no finite expectation and variance (i.e., the respective integrals diverge).

The goal of the present paper is to develop statistical inference for the standard error of inverse prediction (SEIP) in the framework of simple linear regression (1) with normally distributed errors.

The motivation of our metric is as follows. Let $\hat{x}_i$ be the prediction of x given $\hat{y} = y_i$ from formula (3). Then the quality of the dose–response relationship can be quantified by the SEIP as

$$\begin{aligned} SEIP &= \sqrt{\frac{1}{n-2} \sum_{i=1}^n (\hat{x}_i - x_i)^2} = \sqrt{\frac{1}{n-2} \sum_{i=1}^n \left(\frac{y_i - a}{b} - x_i \right)^2} \\ &= \frac{\sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{b} \end{aligned}$$

The choice of $n - 2$ is explained by the fact that $\hat{\sigma}^2$ is an unbiased estimator of σ^2 . This representation justifies the following definition of the empirical SEIP for inverse prediction, denoted $\hat{\eta}$, as

$$\hat{\eta} = \frac{s}{b}, \quad (4)$$

where

$$s = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

the estimate of the regression standard error σ . Obviously, the theoretical SEIP is defined as

$$\eta = \frac{\sigma}{\beta}. \quad (5)$$

As in the case of x -prediction (3), the empirical SEIP (4) does not have a finite mean and variance. We emphasize the difference between the direct (SEP) and indirect (SEIP) regression prediction. SEIP is the metric of the dose–response relationship regardless of the dose design points $x_1, x_2, \dots, x_n$ similarly as σ is the metric of the direct prediction as it follows from Equation (2).

Prediction of x given y in the framework of linear regression model, sometimes referred to as linear calibration problem, has a long history (see more detail in Section 4). The emphasis of the present paper is different—to develop a metric (figure of merit) for calibration curve with inverse prediction that reflects the overall quality of calibration. In a way, the SEIP is more fundamental than individual prediction (3) because it allows comparison between regressions with different slopes, whereas individual predictions assume the same model (1). Consequently, the SEIP can be used for quality comparison across models with different dependent variables and different slopes in the way the signal-to-noise ratio is used in signal processing in technological applications.

From a statistical perspective, SEIP is a generalization of the coefficient of variation (CV) defined as the ratio of the standard deviation to the mean when observations y_i are independent and identically normally distributed as $\mathcal{N}(\mu, \sigma^2)$. Statistical inference, interval estimation, and distribution of CV has a long history going back to papers by McKay [4], Pearson [5], Fieller [6], and Hendricks and Robey [7]. There is evidence of recent interest in CV as well: asymptotic methods for CV are discussed in Miller [8] and Vangel [9] and more recently in Mahmoudvand and Hassani [10] and in the multiple sample framework in Nairy and Rao [11].

The aims of this paper are to (1) generalize CV to include the concept of SEIP and (2) develop new approximate and exact confidence intervals (CIs) with coverage probability close to the nominal even for a small sample size.

2. Confidence intervals for standard error of inverse prediction

The goal of this section is to develop two-sided and one-sided CIs for the true SEIP η using its empirical counterpart $\hat{\eta}$. We start with the delta method to approximate the standard error

of SEIP (SE SEIP) based on the first-order approximation. Then we propose a more accurate CI by observing that the distribution of reciprocal $\hat{\eta}$ follows the noncentral t -distribution. First, we construct a CI based on the normal approximation to the noncentral t -distribution, and second, we construct the exact CI, that is, the CI with the exact coverage probability. In our derivations, we assume that the slope b (and consequently $\hat{\eta}$) is positive. In general cases, we use the absolute value of b , but we assume that with probability close to 1, b is either positive or negative. This assumption holds when the standard error of b is fairly small, so that the CI does not contain zero. This is a fair assumption in many practical situations. At the end of this section, we compare the coverage probability of these CIs via Monte Carlo simulations.

2.1. The delta method for standard error of inverse prediction

Although $\hat{\eta}$ does not have the finite mean and variance, the delta method produces a fairly satisfactory CI for η as follows from our simulations; see details in Section 2.2.3. The following facts are used:

$$\text{var}(s^2) = \frac{2\sigma^4}{n-2}, \quad \text{var}(b) = \frac{\sigma^2}{S_x^2},$$

where

$$S_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

is a fixed number. Using the preceding formulas and because of the independence of s and b , from the delta method, we obtain

$$\text{var}(\hat{\eta}) \simeq \frac{s^2}{b^4} \text{var}(b) + \frac{\text{var}(s)}{b^2} \simeq \frac{\hat{\eta}^2}{2(n-2)} + \frac{\hat{\eta}^4}{S_x^2}.$$

Thus, SE SEIP can be approximated as

$$\text{SE}(\hat{\eta}) \simeq \hat{\eta} \sqrt{\frac{1}{2(n-2)} + \frac{\hat{\eta}^2}{S_x^2}}. \quad (6)$$

On the basis of this formula, an approximate *symmetric* 100(1 - α)% CI for η is

$$\hat{\eta} \pm Z_{1-\alpha/2} \text{SE}(\hat{\eta}), \quad (7)$$

where $Z_{1-\alpha/2}$ is the (1 - $\alpha/2$)th quantile of the standard normal distribution. This CI may be viewed as a generalization of the asymptotic CI derived for the CV in the independent and identically distributed (i.i.d.) case [8, 10, 12].

2.2. Confidence intervals for standard error of inverse prediction

It is more convenient to work with reciprocal SEIP when it comes to CI or hypothesis testing for η . Indeed, whereas $\hat{\eta}$ has neither mean nor variance, the reciprocal has finite mean and variance. Moreover, statistic

$$X = \frac{S_x}{\hat{\eta}} \sim t(\nu, \delta) \quad (8)$$

has the noncentral t -distribution with $\nu = n - 2$ degrees of freedom (DOF) and noncentrality parameter (NCP) $\delta = S_x/\eta$. This follows from the fact that $S_x/\sigma \sim \mathcal{N}(S_x\beta/\sigma, 1)$ and $(n-2)s^2/\sigma^2 \sim \chi^2(n-2)$. For a special case when y_i are i.i.d., this fact was reported in [13, p. 193]. McKay [4] and later Vangel [9] used an approximation in the form of the chi-square distribution to derive a CI for the normal i.i.d. case as well. Our main idea is to construct the CI for $1/\eta$ and then invert it for η . In the next section, we derive CI for SEIP using the normal approximation to the noncentral t -distribution, and next we develop more sophisticated CI with exact coverage probability.

2.2.1. Normal-approximation confidence interval—The easiest way to construct a CI for η is to apply the following normal approximation to the cumulative distribution function (CDF) of the noncentral t -distribution T with DOF = ν and NCP = δ :

$$T(x; \nu, \delta) \simeq \Phi\left(\frac{x - \delta}{\sqrt{1 + \frac{x^2}{2\nu}}}\right), \quad (9)$$

where Φ denotes the CDF of the standard normal distribution [14]. Note that this normal approximation does not use DOF and is also poor for extreme values of x because the right-hand side does not approach either 1 or 0 when $x \rightarrow \pm\infty$. We obtain the approximate equal-tail two-sided CI for η by inverting the inequality

$$\left| \frac{X - \frac{S_x}{\eta}}{\sqrt{1 + \frac{X^2}{2(n-2)}}} \right| < Z_{1-\alpha/2},$$

which yields

$$\frac{\hat{\eta}S_x}{S_x + Z_{1-\alpha/2} \sqrt{\hat{\eta}^2 + \frac{S_x^2}{2(n-2)}}} < \eta < \frac{\hat{\eta}S_x}{S_x - Z_{1-\alpha/2} \sqrt{\hat{\eta}^2 + \frac{S_x^2}{2(n-2)}}} \quad (10)$$

with approximate coverage probability $1 - \alpha$. If the denominator in the upper limit is nonpositive, we set the upper limit to infinity. The one-sided CI for η is obvious to obtain using this normal approximation.

2.2.2. Exact confidence interval—The method for CI with the exact coverage probability based on the inverse CDF is known for a long time. Let statistic X have CDF $F(x; \theta)$, where θ is a one-dimensional parameter. If $F(x; \theta)$, as a function of θ , is strictly decreasing for each x , the lower and the upper limits of the inverse CDF, θ_L and θ_U , are the solutions of the equations $F(X; \theta_L) = 1 - \alpha/2$ and $F(X; \theta_U) = \alpha/2$, respectively, where $1 -$

α is the confidence level. If $F(x; \theta)$ is a strictly increasing function, the lower and the upper limit equations interchange, namely, $F(X; \theta_U) = 1 - \alpha/2$ and $F(X; \theta_L) = \alpha/2$. By convention, we let $\theta_L = -\infty$ and $\theta_U = +\infty$ if the respective equations do not have a solution. The interval (θ_L, θ_U) covers the true value θ with probability $1 - \alpha$. The proof of this statement is found in [15], which refers to [16] as the originator of this method. Indeed, this method was known long before and particularly used by David [17] to construct the exact CI for the Pearson correlation coefficient. We shall refer to this method shortly as the *inverse CDF*.

Using distribution (8), we derive the exact CI for η as the solution to the following equations:

$$T(X; v=n-2, \delta=S_x/\eta_U)=1-\alpha/2, \quad T(X; v=n-2, \delta=S_x/\eta_L)=\alpha/2, \quad (11)$$

where T is the CDF of the noncentral t -distribution and $X = S_x/\hat{\eta}$. The CDF of the noncentral t -distribution is readily available in modern statistical packages including R(function pt). We can solve the preceding equations by a variety of built-in functions, such as uniroot or nls in R. Alternatively, one can compute the left-hand sides of (11) on the grid of values of η and pick those closest to the right-hand side.

In a special case when y_i are i.i.d. normally distributed, Lehmann and Romano [13, p. 224] prove that the one-sided test for μ/σ based on the noncentral t -distribution leads to the uniformly most powerful invariant test.

2.2.3. Comparison of confidence intervals for standard error of inverse prediction via simulation—We compare the coverage probability of three CIs for SEIP (η) via $N = 50,000$ simulations using the true slope $\beta = 1$, the intercept $\alpha = 2$, and the sample size $n = 10$ with $x = 1, 2, \dots, 10$ in the dose–response relationship governed by Equation (1). The regression σ varied from 0.5 to 2.5, which corresponds to the true SEIP, $\eta = 0.5$ to 2.5 with the increment 0.5; see Figure 1. The delta method produces symmetric CI with the SE SEIP approximated by (6). The normal-approximation CI is given by formula (10). The lower and the upper limits of the noncentral t -distribution method are the solutions of Equation (11).

As seen from Figure 1, the delta method consistently underestimates the nominal confidence level. The coverage probability of the normal CI is very close to the nominal 95% but abruptly drops when SEIP approaches 2.5. To the contrary, the coverage probability of the CI based on the noncentral t -distribution remains very close to the nominal value over the entire range of the true SEIP.

3. Hypothesis testing and power computation

There is a well-known duality between CI and hypothesis testing [15, 18, 19]. For example, if the two-sided null hypothesis about an unknown parameter is formulated as $H_0: \eta = \eta_0$ with the alter-native $H_A: \eta \neq \eta_0$, the hypothesis is rejected if the CI does not cover η_0 . The two-sided CI with the exact confidence level will lead to a statistical test with the exact type I error. This link provides the method of testing statistical hypothesis via CI. If the null hypothesis is formulated as $H_0: \eta = \eta_0$ versus $H_A: \eta > \eta_0$, the one-sided CI should be used; see more detail in Section 4.

3.1. Power computation

In this section, we derive the theoretical power function for testing the null hypothesis about SEIP expressed in the form $H_0: \eta = \eta_0$ against the alternative $H_A: \eta \neq \eta_0$ using the normal

approximation and the exact noncentral t -distribution; we validate these tests through statistical simulations.

Suppose that a calibration curve is estimated at design points $x_1, \dots, x_n$ with k repeated measurements of y at each x_i . According to the normal approximation under the null hypothesis

$$\frac{1/\hat{\eta}-1/\eta_0}{\sqrt{V(\hat{\eta}, n, k)}} \simeq \mathcal{N}(0, 1),$$

where $\hat{\eta} = \hat{\sigma}^2/b$ is the estimated SEIP and

$$V(\eta, n, k) = \frac{1}{kS_x^2} + \frac{1}{2(kn-2)\eta^2}$$

is the variance dependent on the alternative SEIP (η), the number of design points (n), the number of repeated measurements (k), and $S_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$ is the sum of squares over distinct values of x . The null hypothesis is rejected if $|Z| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ th normal quantile. Therefore, the power of the normal-approximation test is given by

$$P_N(\eta, n, k) = \Phi\left(-Z_{1-\alpha/2} + \frac{1/\eta - 1/\eta_0}{\sqrt{V(\eta, n, k)}}\right) + 1 - \Phi\left(Z_{1-\alpha/2} + \frac{1/\eta - 1/\eta_0}{\sqrt{V(\eta, n, k)}}\right).$$

Now we derive the power of the exact test based on the noncentral t -distribution. Let X_1 and X_2 be the α th and the $(1 - \alpha/2)$ th quantiles of the noncentral t -distribution with DOF = $k(n - 2)$ and NCP = $\sqrt{k}S_x/\eta_0$. The null hypothesis is rejected if either of the two inequalities occur

$$\begin{aligned} T(S_x/\hat{\eta}; v=kn-2, \delta=\sqrt{k}S_x/\eta_0) &> \alpha/2, \\ T(S_x/\hat{\eta}; v=kn-2, \delta=\sqrt{k}S_x/\eta_0) &> 1-\alpha/2. \end{aligned}$$

Thus, the power of the exact test is given by

$$P_E(\eta, n, k) = T(X_1; v=kn-2, \delta=\sqrt{k}S_x/\eta) + 1 - T(X_2; v=kn-2, \delta=\sqrt{k}S_x/\eta).$$

These powers may be used to determine the number of repeated measurements (k) to achieve a statistically significant hypothesis testing on the target and the observed SEIP.

We compare the powers of the two tests for different number of repeated measurements (k) and verify them using statistical simulations. In Figure 2, we display the powers P_N and P_E functions of the alternative SEIP η with the null SEIP $\eta_0 = 0.5$. The design points are $x = 0, 2, 5, 10, 15$ with $n = 5$. For each η , we ran 10,000 simulations, and we derived empirical power. As follows from this graph, neither normal approximation nor the test based on this

approximation matches well the exact test for $k = 1$, $n = 5$. The bump in the theoretical noncentral t -distribution power points out to some deficiency in the numerical implementation in R.

In Figure 3, we repeat the analysis with $k = 3$ repeated measurements (the actual number of observations is $k n = 15$). The normal-approximation power is much closer to the exact one but still quite deviates for SEIP apart from the null value. On the other hand, simulated/empirical powers are quite close.

In summary, we suggest using the exact test with the noncentral distribution when computing power and determining the required number of repeated measurements to achieve statistical significance.

3.2. Testing the standard error of inverse prediction equality

Assume we want to compare the accuracy of two dose–response relationships in terms of their inverse prediction. This task reduces to the comparison of their SEIPs with the null hypothesis $H_0: \eta_1 = \eta_2$ versus the alternative $H_A: \eta_1 \neq \eta_2$. Assuming that the two samples are independent on the basis of approximation (9), we have

$$Z = \frac{\frac{X_1}{S_{x1}} - \frac{X_2}{S_{x2}}}{\sqrt{\frac{\frac{X_1^2}{S_{x1}^2}}{1 + \frac{X_1^2}{2(n_1-2)}} + \frac{\frac{X_2^2}{S_{x2}^2}}{1 + \frac{X_2^2}{2(n_2-2)}}}} \simeq \mathcal{N}(0, 1) \quad (12)$$

under H_0 , where $X_k = S_{xk}/\hat{\eta}_k$, $k = 1, 2$. Thus, if $|Z| > Z_{1-\alpha/2}$, we reject the hypothesis $H_0: \eta_1 = \eta_2$. Equivalently, we can express the Z -statistic in terms of the empirical SEIPs $\hat{\eta}_1$ and $\hat{\eta}_2$ as follows:

$$Z = \frac{\frac{1}{\hat{\eta}_1} - \frac{1}{\hat{\eta}_2}}{\sqrt{\frac{1}{S_{x1}^2} + \frac{1}{S_{x2}^2} + \frac{1}{2\hat{\eta}_1^2(n_1-2)} + \frac{1}{2\hat{\eta}_2^2(n_2-2)}}}. \quad (13)$$

4. Confidence interval for the individual inverse prediction

In this section, we discuss CI and hypothesis testing for *individual* inverse prediction (3), where y is the individual measurement outside of the calibration data set satisfying the dose–response relationship, namely, $y = \alpha + \beta x + \varepsilon$, where x is the unknown true dose, subject to interval estimation, and ε is the normally distributed measurement error with zero mean and variance σ^2 . As was mentioned earlier, the dose estimate is given by (3). But it is worthwhile to be reminded that we cannot talk about unbiasedness or variance of $\hat{x}$ as an estimator of x because they do not exist (in fact, they are infinite). However, the problem of interval estimation is still valid: for example, we can talk about the CI for the dose (x_L, x_U) , which covers the true value x with a given probability $1 - \alpha$.

As was mentioned in Section 1, the problem of individual prediction with linear regression model has a long history going back to the paper by Eisenhart [20]. Several authors, including Berkson [21], Brown [22], and Buonaccorsi [23], examined statistical properties of $\hat{x}$.

The easiest way to construct a CI for x is to inverse the CI for y . Indeed, as we know, the probability that y is within the interval

$$a+bx \pm q_{1-\alpha/2} s \sqrt{\frac{n+1}{n} + \frac{(x-\bar{x})^2}{S_x^2}} \quad (14)$$

is $1 - \alpha$, where $q_{1-\alpha/2}$ is the $(1 - \alpha/2)$ th quantile of the t -distribution with $n - 2$ DOF, so solving for x , we obtain CI for x [24, pp. 275–283; 25, pp. 22–26]. There exists an alternative approach based on the least squares estimate in regression of x on y , originally suggested by Krutchkoff [3] and further studied by Williams [26], Chow and Shao [27], Brown [25], and Srivastava [28], among others (more discussion is found in Section 6).

Note that (14) uses the equal t -tail symmetrical critical values, $q_{1-\alpha/2} = -q_{\alpha/2}$; in the next section, we consider asymmetric values. Denoting $r = (y - \bar{y})/b$ and using the formula for the quadratic equation, it is elementary to show that the lower and the upper confidence limits for x are [2, p. 84]:

$$x_L = \bar{x} + \frac{r - \sqrt{D}}{H}, \quad x_U = \bar{x} + \frac{r + \sqrt{D}}{H}, \quad (15)$$

where, for the brevity of the presentation, we denote $g = (q_{1-\alpha/2} \hat{\eta})^2$, $H = 1 - gS_x^{-2}$, and $D = r^2 + H[g(n+1)/n - r^2]$, the discriminant of the quadratic equation. We cannot guarantee that the discriminant is nonnegative (usually it is positive); if it is negative, the CI is not defined by this method. The discriminant may be negative only when the empirical SEIP, $\hat{\eta}$, is large, or when the x -values do not spread enough (S_x^2 is small).

From formula (15), we see that the width of CI is small when the SEIP is small and increases with increased SEIP. Naturally, the SEIP is the standard error of the x -prediction and as such may be used as a crude standard error of the dose estimate, $\hat{x}$. This fact may be used for a ‘quick-and-dirty’ CI for $\hat{x}$ as it formally follows from (14): when n is large, the expression under the square root is close to 1 because $(x - \bar{x})^2$ is much smaller than S_x^2 . Ignoring the square root and recalling that $\text{SEIP} = s/b$, we can obtain an approximate $(1 - \alpha)100\%$ symmetric CI as

$$\hat{x} \pm q_{1-\alpha/2} \times \text{SEIP}. \quad (16)$$

Confidence intervals (15) and (16) are two-sided. Sometimes we are interested in the lower limit only. In the example of the radiation dose estimation, our concern is that a person received a dose greater than a certain value, say, 2 Gy. Then the CI is one-sided and formally takes the form (x_L, ∞) , where the lower limit x_L (minimum dose) is computed by the same formula as in (15) but with $q_{1-\alpha/2}$ replaced by $q_{1-\alpha}$ with the negative sign. Then the interval (x_L, ∞) will cover the true radiation dose x with probability $1 - \alpha$.

4.1. Statistical simulation

In this section, we validate the two CIs for individual prediction using simulations ($N = 100,000$) with x -values: 0, 1, 2, 5, 10, each repeated two times ($k = 2$) with the true intercept $\alpha = 0.01$ and the true slope $\beta = 0.02$ (these values are close to the real-life estimates from our example in the next section). Our goal is to understand how the coverage probability is close to the nominal level for different values of the true SEIP. The true SEIP ($\eta = \sigma/\beta$) ranged from 0.25 to 2.5, so the range of σ is 0.25×0.02 to 2×0.02 ; see Figure 4. We generated 100,000 ten-dimensional normally distributed vectors $\mathbf{y}$ for each SEIP with parameters specified earlier. Also we generated 100,000 normally distributed observations $\hat{y}$ that represent individual measurements with the mean $\alpha + \beta x$ and SD = σ , where $x = 2$ is the

true value to predict from y and estimated intercept and slope as $\hat{x} = (y - a)/b$. The first CI is computed by formula (15), and the second CI is the naive CI (16) based on the definition of SEIP as the standard error of $\hat{x}$. The first interval has a coverage probability close to nominal, 95% on the entire range of the true SEIP, but the coverage probability of the naive CI is consistently smaller than 0.95. Thus, we conclude that (15) yields CI with coverage probability very close to nominal, whereas (16) may be used for a ‘quick-and-dirty’ assessment.

5. Example: electron paramagnetic resonance tooth dosimetry

The availability of methods for fast and reliable measurement of absorbed ionizing radiation after events involving large numbers of exposed individuals is an essential component for preparing for possible radiation accidents such as occurred in Chernobyl and Fukushima or acts of terrorism involving ionizing radiation [29–31]. As a part of preparedness to assess dose and treat exposed individuals in such events, the Dartmouth Physically Based Biodosimetry Center for Medical Countermeasures Against Radiation (Dart-Dose CMCR) is one of seven CMCRs recently established by the National Institutes of Health. Electron paramagnetic resonance (EPR) *in vivo* measurements of tooth enamel is a promising method for fast and reliable dose estimation and is especially useful when thousands and maybe hundreds of thousands of people may need to be triaged for medical treatment [32]. A recent paper by Fattibene and Gallens [33] provides an overview of EPR dosimetry.

The potential for the use of EPR measurements in teeth for estimating radiation dose is based on the linear relationship between the amplitude of the EPR signal and the radiation dose [34]. The dose estimation, framed as a linear regression problem, is an inverse prediction problem (1), where y is the EPR amplitude and x is the corresponding dose. Rigorous evaluation of the dose–response relationship may play a crucial role in instrumental development to determine improvements of hardware (e.g., the type of resonator, bridge, or magnet; number of scans or repeated tooth measurements). Different designs lead to different calibration curves and variation around the curve. Which alternative is better? We argue that SEIP is the metric to guide such decisions in EPR tooth dosimetry.

Statistics also can play a central role in using EPR individual tooth measurement to assess the dose and its uncertainty for purposes of medical triage. Especially when very large numbers of people have potentially been exposed, it is important to quickly and accurately differentiate people whose exposure is above or below the threshold for treatment. Therefore, it is important to know, in addition to the estimate of the individual radiation dose value, the CI to assess the probability that the dose is above the threshold for medical treatment.

In this section, we illustrate the use of the SEIP with its CI for decision-making, first to improve the instrument by identifying what surfaces of incisors to measure. Second, we quantify the uncertainty of the individual dose estimate by CI given the calibration curve, to be used in medical decision-making to triage people for treatment.

5.1. Deciding to measure front or back surface of incisor teeth

Although EPR measurements can be made on any type of tooth, the use of incisors is especially attractive because these are more easily measured, are less likely to have caries or restorations, and are the first permanent teeth so even children could be measured. Williams *et al.* [35] reported comparisons of calibration curves based on measurements of the front and back surface of irradiated incisor teeth in the mouth model. We use those data to assess which surface incisor yields a more accurate dose estimate, as we know the actual dose given. Following our approach, this question reduces to computation and statistical

comparison of two SEIPs. We show the tooth EPR data with the least squares estimated dose–response relationship for both surfaces in Figure 5. Here b is the slope of the regression line, and s is the regression standard error. We emphasize that we cannot compare the front and back incisor calibration curves using traditional statistical goodness-of-fit measures such as residual sum of squares or coefficient of determination because the dependent variables and slopes are different. Because the quality of inverse prediction is our concern, the SEIP is the parameter by which we decide what side of the tooth should be measured to achieve the most accurate radiation dose estimate.

We show three 95% CIs for SEIPs in Table I. These methods produce consistent results; the normal-approximation and the exact CI are especially close.

The SEIP for the back surface of incisors is larger, but is the difference statistically significant? To answer this question, we used the Z -test (12). The Z -score statistic is 1.77, which gives the p -value = 0.077—the data fail to reject the null hypothesis that the SEIPs are the same.

On the basis of this table and assuming for this illustration that there is no plausible reason to expect the dose estimate to differ on the front and back surfaces, we would conclude that it does not matter which surface is used in estimating dose, and so other considerations—like ease of making measurements—can determine instrument development. However, it would be important to carry out more measurements to determine whether the surfaces may be statistically significantly different in view of close proximity of the calculated p -value to the borderline 0.05.

5.2. Individual dose prediction for medical decision-making to triage for care

In this section, we illustrate individual radiation dose prediction using the calibration curve from [36] shown in Figure 6. We performed the EPR measurements using an isolated molar that was irradiated externally and then inserted into a gap between teeth of a volunteer. We made three measurements over three consecutive days on six irradiated teeth and used the average of the EPR amplitude adjusted for the tooth size for each dose for the calibration. We assume that the amount of enamel will impact the measurement and that size of tooth is a reasonable estimate of amount of enamel. The SEIP = $s/b = 0.66$ Gy, where $a = 0.1096$ is the intercept estimate, $b = 0.0195$ is the slope estimate, and $s = 0.0129$ is the regression standard error. We computed the 75% CI (0.49, 1.20) for the true SEIP using the normal approximation.

Now we apply this calibration curve for individual dose estimation. Let the EPR amplitude adjusted for the molar tooth size of individual X be 0.16 V. What is his/her radiation dose estimate and 75% CI for the true dose? We show the results in Table II. We compute the dose estimate via inverse prediction as $(0.16 - 0.1096)/0.0195 = 2.59$ Gy. We compute the 75% SEIP-based CI for the true dose on the basis of Equation (16), which yields $2.59 \pm 1.34 \times 0.66$, where $1.34 = q_{1-0.05/2}$, the quantile of the central t -distribution with $n - 2 = 6 - 2 = 4$ DOF. We compute the inverse CI for individual dose prediction on the basis of Equation (15). As expected, the SEIP-based CI underestimates the uncertainty of the dose estimate (the interval is slightly more narrow). We can assess the probability that individual X received radiation more than 2 Gy as $1 - T((2 - 2.59)/0.66, \text{DOF} = 4) = 0.79$, where T is the CDF of the central t -distribution. In this case, we have a strong evidence that the person should be triaged for medical treatment.

6. Discussion and summary points

We have developed a novel figure of merit for calibration curve, called the SEIP. Whereas the quality of direct prediction in the classical regression analysis is expressed by the standard error of the fit, the quality of prediction in inverse regression is expressed by the ratio of the error of the fit to regression slope.

We can view the SEIP metric as a generalization of the coefficient of variation, $CV = \sigma/\mu$, an established statistical parameter that reflects the relative variation. Consequently, statistical inference developed in this paper for SEIP may be interpreted as a generalization of statistical methods previously developed for CV in the case of i.i.d. observations to regression. Whereas traditional (direct) prediction leads to the best regression fit (minimal residual standard error), the inverse prediction leads to the minimum residual standard error divided by the regression slope (SEIP).

Statistical treatment of inverse prediction requires nontraditional approaches because, even for a simple linear regression, the x -estimate does not have finite mean and variance while the interval estimation of x remains valid. Two statistical problems emerge in the context of inverse prediction: (1) developing a metric (and the associated statistical inference) that reflects the quality of the inverse prediction regardless of the design x -values and (2) developing accurate CIs and hypothesis tests for individual prediction when x is estimated from observed y via inverse regression.

We have developed exact and approximate CIs for the SEIP in simple linear regression with normal errors using the fact that the reciprocal SEIP follows the noncentral t -distribution. We have derived the power functions for the hypothesis test about SEIP for the normal approximation and the exact test based on the noncentral t -distribution, and we have investigated the performance of both approaches via statistical simulations.

Individual prediction, as the ratio of two normally distributed random variables, also requires a non-traditional statistical approach. We have adopted the CI for individual prediction and showed how SEIP may be used for a ‘quick-and-dirty’ interval estimation.

We have illustrated major theoretical developments with the EPR tooth dosimetry for comparison of front and back incisor measurements and individual dose prediction using exact and approximate CIs. Probabilistic assessment of individual dose prediction is crucial for medical treatment in the case of nuclear terrorism that may involve a large number of victims.

Finally, we would like to make some comments on the comparison of the classic approach, that is, regressing y on x , undertaken in this paper, with the alternative approach, that is, regressing x on y , called the inverse estimator approach. In the latter case, with the slope and the intercept estimated as $b_I = \Sigma(x_i - \bar{x})(y_i - \bar{y})/\Sigma(y_i - \bar{y})^2$ and $a_I = \bar{x} - b_I \bar{y}$, respectively, we calculate the x -estimate in a straightforward manner as $a_I + b_I y$. As was mentioned earlier, this method was originated by Krutchkoff [3] and was developed further by subsequent authors. Statistical properties of the inverse estimator have been studied by Williams [26] and Halperin [37], among others. Several authors offer a detailed comparison of the two approaches [38–41]. The main conclusion was that the classic approach is difficult to implement for statistical inference, such as hypothesis testing and CI estimation, because the mean square error (MSE) is infinite (does not exist). Indeed, as we mentioned at the beginning of the paper, the mean and variance of the predicted value (3) do not exist, and therefore a nontraditional statistical treatment is required. We argue that MSE is an inappropriate quality measure in the case on inverse prediction. What we actually need in practical applications are CIs, hypothesis testing, and power function for experiment

planning. We have developed these techniques in this paper in the framework of the classic inverse prediction based on the noncentral t -distribution bypassing MSE. Because none of the exact statistical inference with small n (it is usually the case) is available for inverse estimation approach, including the distribution of statistic $\sum_{i=1}^n (x_i - a_i - b_i y_i)^2$, the classic approach simply becomes more practical.

Acknowledgments

This work was supported by two funding agencies in the US Department of Health and Human Services: (1) The National Institutes of Health (NIAID, the Center for Radiation Dosimetry, Grant # U19AI091173) and (2) The Biomedical Advanced Research and Development Authority (BARDA) in the Office of the Assistant Secretary for Preparedness and Response (Contract no. HHSO100201100024C). Dr. Demidenko was also supported by NIH/NCI grants U54 CA151662 and CA130880.

References

1. Neter, J.; Wasserman, W.; Kutner, MH. Applied Linear Statistical Models. 2. IRWIN; Homewood: 1985.
2. Draper, NR.; Smith, H. Applied Regression Analysis. 3. Wiley; New York: 1998.
3. Krutchkoff RG. Classical and inverse regression methods of calibration. *Technometrics*. 1967; 9(3): 425–439.10.2307/1266511
4. McKay AT. Distribution of the coefficient of variation and the extended t -distribution. *Journal of the Royal Statistical Society*. 1932; 95(4):695–698.10.2307/2342041
5. Pearson ES. Comparison of A. T. McKay approximation with experimental sampling results. *Journal of the Royal Statistical Society*. 1932; 95(4):703–704.10.2307/2342043
6. Fieller EC. A numerical test of the adequacy of A. T. McKay's approximation. *Journal of the Royal Statistical Society*. 1932; 95(4):699–702.10.2307/2342042
7. Hendricks WA, Robey KW. The sampling distribution of the coefficient of variation. *Annals of Mathematical Statistics*. 1936; 7:129–144.10.1214/aoms/1177732503
8. Miller EG. Asymptotic test statistics for coefficient of variation. *Communications in Statistics. Simulation and Computation*. 1991; 20(10):3351–3363.10.1080/03610929108830707
9. Vangel MG. Confidence intervals for a normal coefficient of variation. *The American Statistician*. 1996; 50(1):21–26.10.2307/2685039
10. Mahmoudvand R, Hassani H. Two new confidence intervals for the coefficient of variation in a normal distribution. *Journal of Applied Statistics*. 2009; 36(4):429–442.10.1080/02664760802474249
11. Nairy KS, Rao KA. Tests of coefficients of variation of normal population. *Communications in Statistics. Simulation and Computation*. 2003; 32(3):641–661.10.1081/SAC-120017854
12. Curto JD, Pinto JC. The coefficient of variation asymptotic distribution in the case of non-iid random variables. *Journal of Applied Statistics*. 2009; 36(1):21–32.10.1080/02664760802382491
13. Lehmann, EL.; Romano, JP. Testing Statistical Hypotheses. 3. Springer; New York: 2005.
14. Johnson NL, Welch BL. Applications of the noncentral t -distribution. *Biometrika*. 1940; 31(3/4): 362–389.10.2307/2332616
15. Casella, G.; Berger, RL. Statistical Inference. Duxbury Press; Belmont, CA: 1990.
16. Mood, AM.; Graybill, FA.; Boes, DC. Introduction to the Theory of Statistics. 3. McGraw-Hill; New York: 1974.
17. David, FN. Tables of the Distribution of the Correlation Coefficient. Biometrika Office; London: 1938.
18. Lindley DV, East DA, Hamilton PA. Tables for making inferences about the variance of a normal distribution. *Biometrika*. 1960; 47(3–4):433–435.10.1093/biomet/47.3-4.433
19. Rao, CR. Linear Statistical Methods and Its Applications. 2. Wiley; New York: 1973.

20. Eisenhart C. The interpretation of certain regression models and their use in biological and industrial research. *Annals of Mathematical Statistics*. 1939; 10:162–186.10.1214/aoms/1177732214
21. Berkson J. Estimation of linear function for a calibration line: consideration of a recent proposal. *Technometrics*. 1969; 11(4):649–660.10.2307/1266889
22. Brown PJ. Multivariate calibration. *Journal of the Royal Statistical Society, Series B*. 1982; 44(3):287–321.
23. Buonaccorsi JP. Design considerations for calibration. *Technometrics*. 1986; 28(2):149–155.10.2307/1270451
24. Graybill, FA. *Theory and Application of the Linear Model*. Duxbury Press; North Scituate: 1976.
25. Brown, PJ. *Measurement, Regression, and Calibration*. Clarendon Press; Oxford: 1993.
26. Williams EJ. A note on regression methods in calibration. *Technometrics*. 1969; 11(1):189–192.10.2307/1266774
27. Chow S-C, Shao J. On the difference between the classical and inverse methods of calibration. *Journal of the Royal Statistical Society, Series C*. 1990; 39(2):219–228.10.2307/2347761
28. Srivastava MS. Comparison of the inverse and classical estimators in multi-univariate linear calibration. *Communications in Statistics-Theory and Methods*. 1995; 24(11):2753–2767.10.1080/03610929508831647
29. Swartz HM, Flood AB, Gougelet RM, Rea ME, Nicolalde RJ, Williams BB. Critical assessment of biodosimetry methods for large-scale incidents. *Health Physics*. 2010; 98(2):95–108.10.1097/HP.0b013e3181b8cfff [PubMed: 20065671]
30. Flood AB, Nicolalde RJ, Demidenko E, Williams BB, Shapiro A, Wiley AL, Swartz HM. A framework for comparative evaluation of dosimetric methods to triage a large population following a radiological event. *Radiation Measurements*. 2011; 46(9):916–922.10.1016/j.radmeas [PubMed: 21949481]
31. Swartz HM, Williams BB, Nicolalde RJ, Demidenko E, Flood AB. Overview of biodosimetry for management of unplanned exposures to ionizing radiation. *Radiation Measurements*. 2011; 46(9):742–748.10.1016/j.radmeas
32. Swartz HM, Burke G, Coey M, Demidenko E, Dong R, Grinberg O, Hilton J, Iwasaki A, Lesniewski P, Kmiec M, Lo KM, Nicolalde RJ, Ruuge A, Sakata Y, Sucheta A, Walczak T, Williams BB, Mitchell CA, Romanyukha A, Schauer DA. In vivo EPR for dosimetry. *Radiation Measurements*. 2007; 42(6–7):1075–1084.10.1016/j.radmeas [PubMed: 18591988]
33. Fattibene P, Gallens F. EPR dosimetry with tooth enamel: a review. *Applied Radiation and Isotopes*. 2010; 68(11):2033–2116.10.1016/j.apradiso.2010.05.016 [PubMed: 20599388]
34. Swartz HM. Long-lived electron spin resonances in rats irradiated at room temperature. *Radiation Research*. 1965; 24(4):579–586.10.2307/3571876 [PubMed: 14275318]
35. Williams BB, Dong RH, Nicolalde RJ, Matthews TP, Gladstone DJ, Demidenko E, Zaki BI, Salikhov IK, Lesniewski PN, Swartz HM. Physically-based biodosimetry using in vivo EPR of teeth in patients undergoing total body irradiation. *International Journal of Radiation Biology*. 2011; 87(8):766–775.10.3109/09553002.2011.583316 [PubMed: 21696339]
36. Demidenko E, Williams BB, Sucheta A, Dong R, Swartz HM. Radiation dose reconstruction from L-band in vivo EPR spectroscopy of intact teeth: comparison of methods. *Radiation Measurements*. 2007; 42(6–7):1089–1093.10.1016/j.radmeas.2007.05.025 [PubMed: 18591987]
37. Halperin M. On inverse estimation in linear regression. *Technometrics*. 1970; 12(4):727–736.10.2307/1267319
38. Osborne C. Statistical calibration: a review. *International Statistical Review*. 1991; 59(3):309–336.10.2307/1403690
39. Tellinghuisen J. Inverse vs. classical calibration for small data sets. *Fresenius Journal of Analytical Chemistry*. 2000; 368(6):585–588.10.1007/s002160000556 [PubMed: 11228707]
40. Ali MA, Ashkar MY. The calibration problem revisited. *Communications in Statistics-Theory and Methods*. 2002; 31(10):1733–1741.10.1081/STA-120014911
41. Kannan N, Keating JP, Mason RL. A comparison of classical and inverse estimators in the calibration problem. *Communications in Statistics-Theory and Methods*. 2007; 36(1–4):83–95.10.1080/03610920600966225

Figure 1.

The coverage probability of three confidence intervals as a function of the true SEIP, η . The number of simulations is 50,000, and the nominal confidence level is 95%. SEIP, standard error of inverse prediction.

Figure 2.

Two theoretical and empirical power functions for $k = 1$ (no repeated measurements) and $n = 5$. The normal approximation does not match well the exact test especially for large alternative values of SEIP. SEIP, standard error of inverse prediction.

Figure 3.

Theoretical and empirical power functions for three repeated measurements, $k = 3$ (the total number of measurements is 15). SEIP, standard error of inverse prediction.

Figure 4. Statistical simulations with two CIs for individual prediction based on $N = 100,000$ simulation experiments (the sample size, $n = 10$). CI, confidence interval; SEP, standard error of prediction.

Figure 5. Two calibration curves from front and back incisor surfaces in the mouth model experiments. Each point represents an independent EPR measurement (three different teeth per dose). EPR, electron paramagnetic resonance; SEIP, standard error of inverse prediction.

Figure 6.

Tooth-size adjusted EPR signal amplitude as a function of the radiation dose received with the 75% confidence band (shaded area). If the tooth-size adjusted amplitude is 0.16, the dose estimate is 2.59 Gy with the 75% CI (1.57,3.58) as the minimum and the maximum values on the x -axis of intersection, the horizontal line with the ordinate 0.16 with the shaded area. CI, confidence interval; EPR, electron paramagnetic resonance; SEIP, standard error of inverse prediction.

Table I

SEIP for front and back incisors and their 95% confidence intervals.

Tooth SEIP Method	Front incisor 1.09 Gy		Back incisor 1.84 Gy	
	Lower	Upper	Lower	Upper
Delta (7)	0.687	1.500	1.151	2.519
Normal approximation (10)	0.744	1.618	1.193	2.875
Exact noncentral <i>t</i> -distribution (11)	0.750	1.633	1.231	2.984

SEIP, standard error of inverse prediction.

Table II

Individual dose prediction with the EPR amplitude = 0.16 V.

Method	Lower	Upper
SEIP: Equation (16)	1.71	3.47
Inverse CI: Equation (15)	1.57	3.58

Confidence level = 75%; dose estimate = 2.59 Gy.