Optimal Calibration in Immunoassay and Inference on the Coefficient of Variation

Johannes Forkman

Faculty of Natural Resources and Agricultural Sciences
Department of Energy and Technology
Uppsala

Optimal Calibration in Immunoassay and Inference on the Coefficient of Variation

Abstract

This thesis examines and develops statistical methods for design and analysis with applications in immunoassay and other analytical techniques. In immunoassay, concentrations of components in clinical samples are measured using antibodies. The responses obtained are related to the concentrations in the samples. The relationship between response and concentration is established by fitting a calibration curve to responses of samples with known concentrations, called calibrators or standards. The concentrations in the clinical samples are estimated, through the calibration curve, by inverse prediction.

The optimal choice of calibrator concentrations is dependent on the true relationship between response and concentration. A locally optimal design is conditioned on a given true relationship. This thesis presents a novel method that accounts for the variation in the true relationships by considering unconditional variances and expected values. For immunoassay, it is suggested that the average coefficient of variation in inverse predictions be minimised.

In immunoassay, the coefficient of variation is the most common measure of variability. Several clinical samples or calibrators may share the same coefficient of variation, although they have different expected values. It is shown here that this phenomenon can be a consequence of a random variation in the dispensed volumes, and that inverse regression is appropriate when the random variation is in concentration rather than in response.

An estimator of a common coefficient of variation that is shared by several clinical samples is proposed, and inferential methods are developed for common coefficients of variation in normally distributed data. These methods are based on McKay's chi-square approximation for the coefficient of variation. This study proves that McKay's approximation is noncentral beta distributed, and that it is asymptotically normal with mean $n - 1$ and variance slightly smaller than $2(n - 1)$.

Keywords: calibration, coefficient of variation, four-parameter logistic function, immunoassay, inverse prediction, inverse regression, McKay's approximation

Author's address: Johannes Forkman, Department of Energy and Technology, SLU
Box 7032, 750 07 Uppsala, Sweden
E-mail: Johannes.Forkman@et.slu.se

Dedication

Till Jenny

As our circle of knowledge expands, so does the circumference of darkness surrounding it.

    Albert Einstein

.

Contents

List of Publications

This thesis is based on the work contained in the following papers, which are referred to in the text by their Roman numerals:

I   Forkman, J. A method for designing nonlinear univariate calibration. *Technometrics*. In press.

II Forkman, J. & Söderström, L. (2008). Calibration in immunoassay with proportional errors. Centre of Biostochastics, Swedish University of Agricultural Sciences. Research Report 2008:10.

III Forkman, J. Estimator and tests for common coefficients of variation in normal distributions. *Communications in Statistics – Theory and Methods*. In press.

IV Forkman, J. & Verrill, S. (2008). The distribution of McKay's approximation for the coefficient of variation. *Statistics & Probability Letters* 78, 10-14.

Paper I is reproduced with the permission of the American Society for Quality and the American Statistical Association, Paper III is reproduced with the permission of Taylor & Francis, and Paper IV is reproduced with the permission of Elsevier.

1 Introduction

This thesis examines methods that are useful for statistical design and analysis of analytical procedures (*i.e.* analytical techniques), especially those with applications within immunoassay. Analytical procedures are techniques used in clinical chemistry for determining concentrations of chemical substances. Immunoassay is an analytical procedure that uses antibodies (Hage, 1999). In other words, the immunoassay is a laboratory test, based on antibodies, that measures the concentration of an analyte in a clinical sample obtained from a patient. Because immunoassay is a complex technique, sample concentrations are usually determined with significant errors, which can be regarded as systematic or random. The standard deviation of the measurements is often proportional to the average of the measurements. For this reason, the coefficient of variation, *i.e.* the standard deviation divided by the mean, is a common measure of variability in immunoassay.

Section 2 clarifies some useful concepts in immunoassay and statistical modelling. As a background to the research, Section 3 includes a review of statistical methods for univariate calibration and for inference on the coefficient of variation. Section 4 presents the results of the thesis, based on Papers I–IV, which are included as appendices. Section 5 provides a list of conclusions, while Section 6 is a summary in Swedish.

1.1 Objectives

In statistical textbooks, variance is the central measure of dispersion and homogeneous variances is often assumed. Various methods are available for statistical inference on the variance. However, in many applications, relative deviations are more interesting than absolute deviations. In immunoassay, the coefficient of variation is the measure of variability most commonly used among researchers. The concern in diagnostic research about relative errors,

rather than absolute errors, was the starting point for this thesis and its theme throughout.

The main aims of this thesis were to develop statistical methods based on the assumption of homogeneous coefficients of variation and with the focus on minimising relative errors, and to contribute to the methods for inference on the coefficient of variation.

The specific objective was to develop statistical methods for calibration and for statistical inference on variation, with applications in immunoassay.

2 Basic Concepts

Before entering into the main topics of the thesis, some important concepts in immunoassay and statistical modelling must be clarified. Section 2.1 provides some basic knowledge about immunoassay. Section 2.2 considers some functions that are often used in immunoassay for modelling the relationship between response and concentration, while Section 2.3 discusses some common assumptions that are often made about the distribution of the random errors in immunoassay measurements.

2.1 Immunoassay

An analyte is the chemical component measured in an analytical procedure. In immunoassay, the analyte is either an antibody or an antigen. Antibodies are proteins in the blood that are produced by the immune system for protection against foreign bodies, while the foreign bodies are the antigens. The antibodies bind to the antigens.

The antigenes or the antibodies are labelled before analysis, in order to give a measurable signal. This label can be an enzyme, a radioactive isotope, or fluorescein. The signals obtained from an immunoassay can be radioactivity or emission of light. These signals are commonly called responses.

The immunoassay involves chemical reactions between clinical samples obtained from patients and reagents (*i.e.* chemical solutions) performed under standardised conditions. The result is a response that is related to the concentration of the analyte in the sample. In competitive immunoassay, the analyte is unlabelled and competes with labelled molecules. The response is then a decreasing function of the analyte concentration. In noncompetitive immunoassay the labelled molecules bind to the analyte, and the response is an increasing function (Figure 2.1). In either case, the exact relationship

between response and concentration needs to be estimated. This estimation is called calibration.
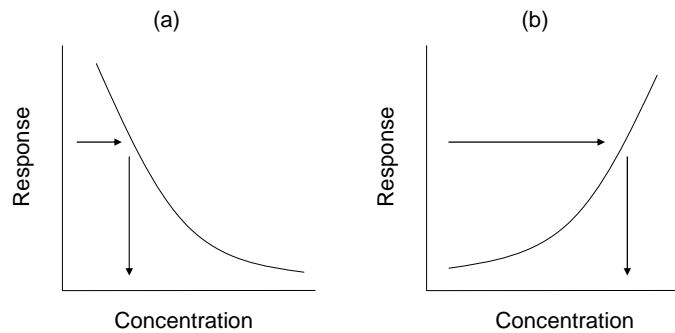

Figure 2.1. a) Competitive immunoassay, b) noncompetitive immunoassay.

For calibration, samples with known concentrations are required. These specific samples are called calibrators or standards, and are usually prepared in advance. For example, a single sample with a known high concentration can be dissolved in water or animal serum to produce calibrators with a few specified lower concentrations covering the range of measurement. When discussing statistical design for calibration, the specified calibrator concentrations are called design points. Because the calibrators are specially prepared, but the samples are not, the calibrators and the clinical samples may react in slightly differently ways.

Usually, a set of clinical samples with unknown concentrations is assayed together with the calibrators in an assay run. A calibration curve is fitted to the responses of the calibrators. This curve can be a straight line or some other monotonic function. The responses of the clinical samples are transformed into estimates of concentration through the fitted calibration curve, as illustrated by the arrows in Figure 2.1. This method for estimation of sample concentrations is called inverse prediction.

Because the relationship between response and concentration may change from one assay run to another, calibrators are often included in each assay run, so that each can be calibrated separately. However, in some systems it is assumed that the relationship is stable, so that calibration needs to be performed less often, for example only once a month or when new batches of reagents are taken into use.

There are many sources of variation in immunoassay results. Small variations in operation, for example in timing, incubation temperature, relative humidity or dispensed volume, can add up to a substantial random error. In controlled precision studies, such sources can be identified as variance components. Usually the variation is estimated between and within assay runs, giving estimates of inter-assay and intra-assay variation, respectively. However, more factors than assay runs can be included in a precision study. In well-performed precision studies it is possible to estimate variances between for example laboratories, instruments, batches of reagents and batches of calibrators.

In a precision profile, the variation in the immunoassay is plotted versus concentration as points or as a continuous curve, as in Figure 2.2. The variation can be the intra-assay variation or the total variation. The standard deviation is usually presented as a percentage of the concentration level (*i.e.* as a coefficient of variation).
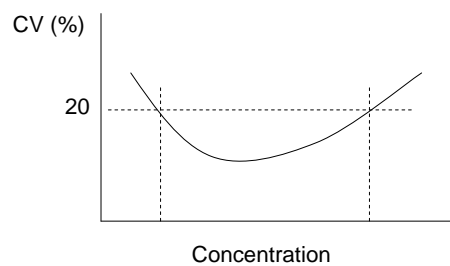
Figure 2.2. Precision profile with indicated working range as defined by Carroll (2003).

The range of measurement of the immunoassay method should begin above the quantitation limit, where samples are 'quantitatively determined with stated acceptable precision and trueness' (Clinical and Laboratory Standards Institute, 2004). Similarly, it should end before the errors in the inverse predictions become too large (Gottschalk & Dunn 2005a). The working range is a related concept, defined for example as 'the range of concentrations for which the coefficient of variation of the fitted concentrations is <0.2.' (Carroll, 2003). In the precision profile, concentrations are often displayed on a logarithmic scale.

2.2 Dose-Response Curves for Calibration

In immunoassay, the response y is usually considered to be the sum of a function f of the concentration x and a random error term e, that is, $y = f(x) + e$.

The straight line, $f(x) = \beta_1 + \beta_2 x$, is often used as a calibration curve in analytical procedures. Polynomials, such as the second-order polynomial $f(x) = \beta_1 + \beta_2 x + \beta_2 x^2$, could fit significantly better, but may not be appropriate, since they need not be monotonic in the range of measurement.

The Michaelis–Menten function (e.g. Wagner, 1973),

$$f(x) = \frac{\beta_2 x}{\beta_1 + x}, \tag{2.1}$$

is a simple monotonic function that is nonlinear in the parameters. This function can sometimes be used for describing the relationship between response and concentration in analytical procedures. It is mainly used in kinetics, and the asymptotic response β_2 is then called V_{max}, to denote the maximum velocity. The Michaelis constant β_1, often denoted K_m, is the concentration x that gives half the maximum response, *i.e.* $f(x) = \beta_2/2$.

The four-parameter logistic function,

$$f(x) = \beta_2 + \frac{\beta_1 - \beta_2}{1 + \left(\dfrac{x}{\beta_3}\right)^{\beta_4}}, \tag{2.2}$$

is commonly used for calibration in immunoassay (O'Connell, Belanger & Haaland, 1993). The parameter β_1 is the response $f(x)$ at concentration $x = 0$, and β_2 is the limit of $f(x)$ when x approaches infinity. The parameter β_3 is often denoted ED50. At concentration ED50, the function takes the value $(\beta_1 + \beta_2)/2$. The slope of the curve is controlled by β_4.

The four-parameter logistic function can be extended to a five-parameter logistic function (Rodbard, Munson & De Lean, 1974), as

$$f(x) = \beta_2 + \frac{\beta_1 - \beta_2}{1 + \left(\dfrac{x}{\beta_3}\right)^{\beta_4} \left(\dfrac{1}{2}\left(1 + \dfrac{x}{\beta_3}\right)\right)^{\beta_5 - \beta_4}}. \tag{2.3}$$

The parameters β_1, β_2 and β_3 in (2.3) have the same interpretations as in (2.2). The parameters β_4 and β_5 in (2.3) control the slope of the curve. Parameter β_4 affects the curve mainly below ED50, while β_5 affects the curve mainly above ED50. There is no explicit expression for the inverse of the five-parameter logistic function. In other words, it is not possible to solve for x in (2.3). Consequently, numerical methods are needed for inverse predictions. For example, the secant method could be used.

Another extension of (2.2) into a logistic function with five parameters is

$$f(x) = \beta_2 + \frac{\beta_1 - \beta_2}{\left(1 + \left(\dfrac{x}{\beta_3}\right)^{\beta_4}\right)^{\beta_5}} . \qquad (2.4)$$

Gottschalk & Dunn (2005b) studied this function and concluded that in immunoassay, it can substantially improve the accuracy of inverse prediction, compared to the four-parameter logistic function (2.2). With (2.4), ED50 could be larger or smaller than β_3, depending on the values of β_4 and β_5. Inverse prediction is easy, because it is possible to solve for x in (2.4). A third extension of (2.2) into a logistic function with five parameters was proposed by Ricketts & Head (1999).

The three-parameter logistic function is obtained as a special case of (2.2) when $\beta_1 = 0$, and the Michaelis–Menten function (2.1) is obtained if, in addition, $\beta_4 = 1$.

2.3 Probability Distributions

It is necessary to know the distribution of the error term, or to make assumptions about it, in order to apply parametric methods for statistical inference. Furthermore, if we want to fit the calibration curve by the method of maximum likelihood, it is necessary to know the distribution.

The error term is commonly assumed to be normally distributed with expected value 0. Many standard statistical techniques, such as analysis of variance and general linear modelling, require normally distributed observations, although they may be robust for minor deviations from normality. Sometimes the central limit theorem, which assumes the sum of many distributed random variables to be approximately normally distributed, is used as a heuristic argument for normally distributed measurements. The responses obtained in immunoassay can indeed be regarded as sums of many random variables, but the conditions of the theorem (e.g. Gut, 2005) are hard to verify. In practice, it can be difficult to claim that all observations are

positively or negatively skewed if some samples are slightly skewed to the left and others to the right. In the absence of arguments for skewed distributions, the normal distribution is often chosen.

When the magnitude of the random error terms is independent of the level of the observations, the data are said to be homoscedastic and the variance is constant. However in immunoassay, the data are often heteroscedastic. This may become apparent in a residual plot, with residuals increasing by predicted values. Assumptions are required about how the variance of the normal distribution is changing. For example, it could be assumed that the variance follows the power-of-the-mean model,

$$\text{var}(y) = \phi\,(\text{E}(y))^{\theta},\tag{2.5}$$

or the log-linear model,

$$\text{var}(y) = \phi\,\exp(\theta x).\tag{2.6}$$

The use of functional relationships such as (2.5) or (2.6), between the variance and the level of the measurements, is recommended, because using sample variances as weights is inefficient when the number of replicates is small (Carroll & Cline, 1988). Methods for estimation of the variance parameters ϕ and θ are briefly discussed in Section 3.2.1.

When the standard deviation is proportional to the averages, *i.e.* when $\theta = 2$ in (2.5), it is convenient to assume that the observations follow a lognormal distribution, rather than a normal distribution. According to this assumption, y is skewed to the right, and $\log y$ is normally distributed. We then assume that $\log y = \log f(x) + \log e,$ and that the observations follow a multiplicative model, *i.e.* $y = f(x)e.$ By a Taylor series expansion of y about the expected value,

$$\log y \approx \log \text{E}(y) + \frac{1}{\text{E}(y)}(y - \text{E}(y))$$

so that $\text{var}(\log y) \approx \text{var}(y)/(\text{E}(y))^{2}$. Thus, the logarithmic transformation is variance-stabilising when the coefficient of variation is constant on the original scale.

The probability density function of the lognormal distribution can be written

16

$$\frac{1}{y\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(\log y-\mu)^2\right\}. \tag{2.7}$$

The expected value of the lognormal distribution is $\exp(\mu+\sigma^2/2)$, and the variance is $\exp(2\mu+\sigma^2)(\exp(\sigma^2)-1)$. Consequently the squared coefficient of variation γ^2 is $\exp(\sigma^2)-1$.

Sometimes the measurements of the immunoassay are assumed to be gamma distributed. The probability density function of the gamma distribution is usually parameterised in terms of a shape parameter κ and a scale parameter λ, as

$$\frac{y^{\kappa-1}\exp(-y/\lambda)}{\Gamma(\kappa)\lambda^{\kappa}}. \tag{2.8}$$

In this parameterisation, the expected value is $\kappa\lambda$, and the variance is $\kappa\lambda^2$. The squared coefficient of variation γ^2 equals $1/\kappa$. The expected value and the coefficient of variation are orthogonal (*i.e.* their information matrix is diagonal), and their maximum likelihood estimators are asymptotically independent (Cox & Reid, 1987). Under the assumption of constant coefficients of variation, the measurements could be modelled by gamma distributions with equal shape parameters, but varying scale parameters.

When the coefficient of variation is small, the differences between the normal distribution, the lognormal distribution and the gamma distribution are small. This is illustrated in Figure 2.3.
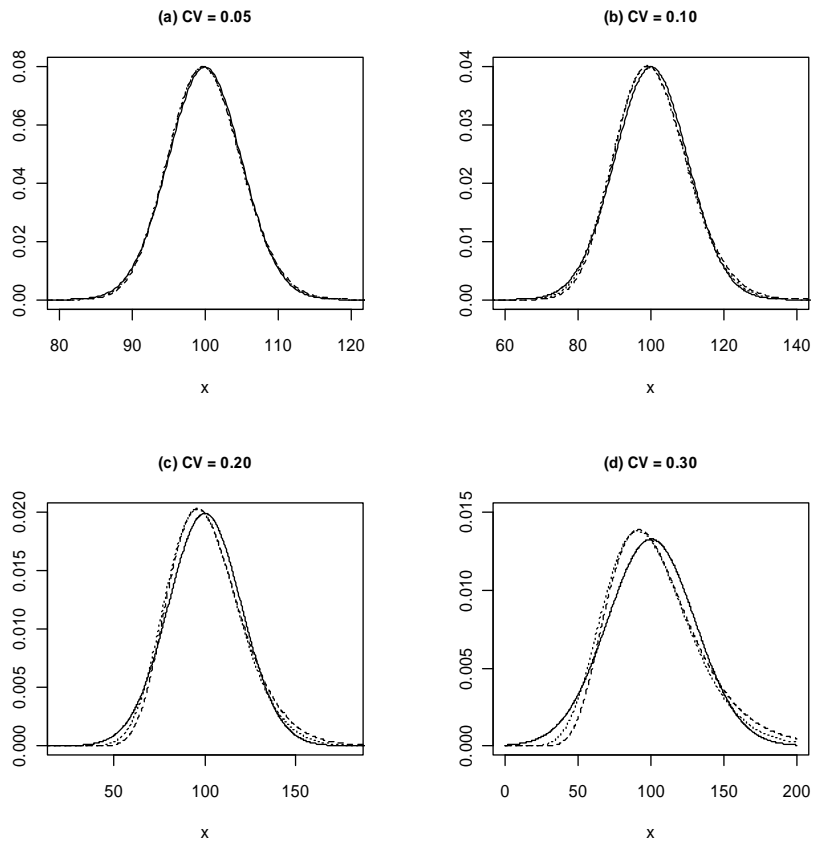
Figure 2.3. Probability density functions of normal distributions (solid), lognormal distributions (dashed) and gamma distributions (dotted). Expected value: 100, Coefficient of variation (CV): (a) 0.05, (b) 0.10, (c) 0.20, (d) 0.30.

3 Background

This section reviews and discusses statistical methods for design of calibration, criteria for calibration, and inference on the coefficient of variation as a background to the research performed.

Calibration design refers to the choice of calibrator concentrations. There are three main types of optimal design: locally optimal designs, Bayesian optimal designs and maximin optimal designs. In addition, many criteria for optimality have been proposed, as reviewed in Section 3.1.

By calibration criteria are meant conditions for the optimal fits of calibration curves. According to the method of least squares, the calibration criterion that should be minimised is the sum of the squared deviations between the obtained and the predicted responses. This criterion is discussed in Section 3.2.1. A similar criterion is used when the curve is fitted by inverse regression. According to this method, the sum of squares is minimised as measured on the concentration axis, in preference to the response axis. This is described in Section 3.2.2. Calibration curves can also be fitted by maximum likelihood or, in presence of heteroscedasticity, by the transform–both–sides method, as described in Section 3.2.3. Numerical methods for fitting of calibration curves are not discussed in this thesis. Seber & Wild (1989) provide information about such techniques.

Many statistical methods have been developed for inference on the coefficient of variation, but in immunoassay they are rarely used. Warren (1982) wrote:

> While workers in many fields recognize the imprecision in a sample mean, and will now routinely compute a standard error, or a confidence interval, for the mean, many of these same workers will treat the sample coefficient of variation as if it were an absolute quantity. Inferences based on this measure of variability may then be questionable. Nevertheless, it should be possible to

persuade such workers that, as with the sample mean, some measure of precision should be attached to the sample coefficient of variation.

Although many years have passed since this reflection was first published, the situation has not improved, perhaps because methods for inference on the coefficient of variation are seldom discussed in statistical textbooks. Section 3.3 provides a review of methods for inference on the coefficient of variation.

3.1 Calibration Design

Consider the curve in Figure 3.1. Five calibrators, with concentrations 2, 10, 50, 100 and 200 µg/l, were used for calibration. The calibrators were measured in an assay run, and a calibration curve was fitted to the responses. We may wonder whether $\xi = (2, 10, 50\ 100, 200)'$ is an optimal design, or whether there exists some other set of design points (*i.e.* calibrator concentrations) that perform better with regard to some design criterion.

A number of design criteria have been proposed. Many of these focus on the variation in the estimators of the curve parameters, as described by a variance-covariance matrix $\mathbf{V}$ that is exact in the linear case and asymptotic in the nonlinear case. Let $\mathbf{D}$ denote the variance-covariance matrix of the calibrators. This matrix includes the intra-assay variances in the responses of the calibrators, and is usually considered to be diagonal because of independent measurements. The diagonal elements of $\mathbf{D}$ are all equal in the case of homoscedasticity and unequal otherwise. When the calibration curve is linear in the parameters and $\mathbf{X}$ is the design matrix, the variance-covariance matrix $\mathbf{V}$ of the parameter estimates equals $(\mathbf{X'D}^{-1}\mathbf{X})^{-1}$. When the calibration curve is nonlinear, the asymptotic variance-covariance matrix $\mathbf{V}$ is $(\mathbf{F'D}^{-1}\mathbf{F})^{-1}$, where $\mathbf{F}$ is a matrix of partial derivatives of the calibration function with respect to the parameters. Under the assumption of independent normally distributed observations, the inverse of $\mathbf{V}$ equals the expected information matrix (e.g. Seber & Wild, 1989, p. 34). In *D*-optimal deigns, the determinant of $\mathbf{V}$ is minimised. In *A*-optimal designs, the trace of $\mathbf{V}$ is minimised, and in *E*-optimal designs, the maximum eigenvalue of $\mathbf{V}$ is minimised. By these three design criteria, different measures of the variation in the parameter estimators are made as small as possible.
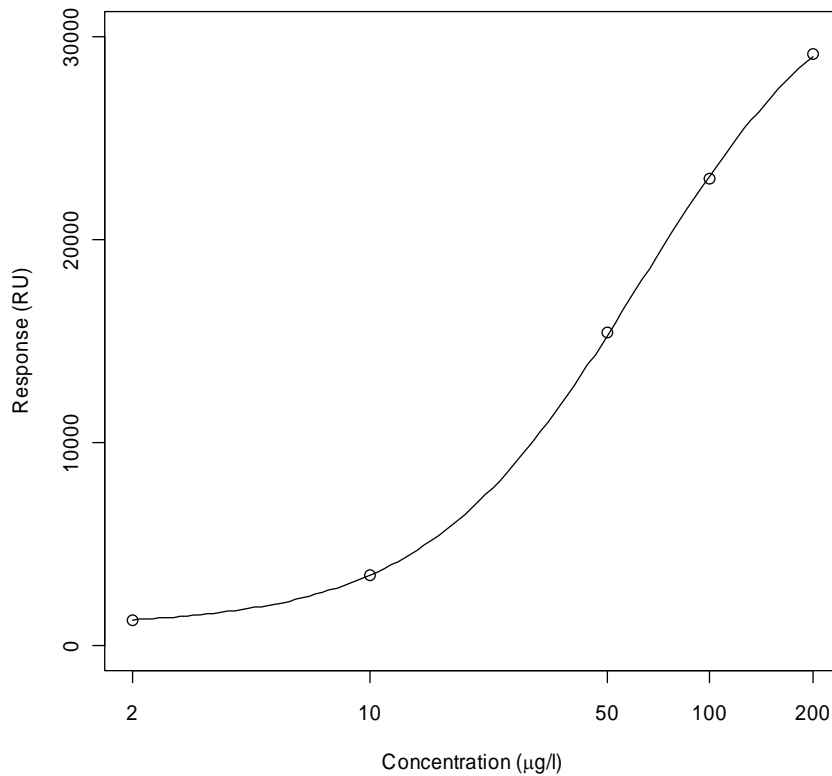
Figure 3.1. A four-parameter logistic calibration curve fitted to responses of five calibrators.

Instead of focusing on the precision in the curve parameter estimates, the focus could be on the precision in the predictions that belong to a specified interval of interest. In *G*-and *V*-optimal designs, the maximum and average variance in the predictions is minimised, respectively. *V*-optimal designs are sometimes called *I*-optimal, because their *integrated* prediction variance is minimised (e.g. Hardin & Sloane, 1993). François, Govaerts & Boulanger (2004) proposed that *G*– or *V*-criteria be used, in the calibration context, for inverse predictions that belong to a range of measurement. Rocke & Jones (1997) similarly proposed maximisation of reciprocals of variances in inverse predictions.

In nonlinear regression, the variance-covariance matrix **V** and the inverse predictions are dependent on the current curve parameter values. This is a problem when searching for an optimal design, because the parameter values change between assay runs, for example as a result of changes in temperature

or reagents. As a consequence, the design that is optimal in one assay run may not be optimal in another.

A locally optimal design is a design that is optimal, according to some design criterion, for a given set of fixed parameter values. It could be argued that it is reasonable to search for a locally optimal design, because 'the researcher often has some knowledge on the expected model parameters' (François, Govaerts & Boulanger, 2004). On the other hand, it is logically apparent that locally optimal designs are not necessarily optimal when the parameter values change.

Bayesian methods provide a solution to this problem. According to this approach the parameter values are modelled as random variables. Bayesian D- or A-optimality, which take into account the prior variation in the parameter values, can be derived (Chaloner & Verdinelli, 1995). Bayesian methods are computer intensive, because they require numerical integration or use of simulation based methods (Müller, 1999). Furthermore, in practice it can be difficult to define the prior distribution of the curve parameters, especially because the parameters may be correlated and have different ranges.

Instead of using Bayesian methods, the researcher can try to find a maximin optimal design by specifying a subset of the parameter space, including all parameter values that should be considered, possibly all values that could occur. In the case of one parameter, the subset is usually an interval, and otherwise usually a product set of intervals. The researcher focuses on a design criterion that should be maximised, and maximises over the design space the minimum of this criterion over the parameter subset. Through this method, designs that perform less well for some possible parameter values are avoided. However, the method does not take into account the likelihood of the possible parameter values. The basic criterion that is maximised according to the maximin method is usually a function of the information matrix, thus giving maximin D-, A- or E-optimal designs. Such designs have been suggested for the Michaelis-Menten model by Dette & Biedermann (2003), Dette, Melas & Pepelyshev (2003) and Dette, Melas & Wong (2005).

Errors-in-variables models (Carroll, Ruppert & Stefanski, 1995) can be considered when there are random errors in the calibrator concentrations. Zwanzig (2000) and Pronzato (2002) have studied design problems for such models.

3.2 Calibration Criteria

By calibration is meant establishment of the current relationship between response and concentration. In practice the calibration is usually made by fitting the calibration curve to the responses of the calibrators. It is assumed that the dose–response function f is known, but the current parameter values are not. Under this assumption, the immunoassay calibration amounts to estimation of the curve parameters. This section discusses criteria for estimation of the vector $\boldsymbol{\beta}$ of curve parameters, in the presence of heteroscedasticity. The function f is regarded as a function of $\boldsymbol{\beta}$ and is written $f(\boldsymbol{\beta})$. The vector of calibrator responses is denoted $\mathbf{y}$.

3.2.1 Weighted least squares

With the method of least squares, the residual sum of squares is minimised. The parameter values that minimise $\mathbf{e'e}$ are taken as curve parameter estimates, where $\mathbf{e}$ denotes the vector of residuals. When the data are heteroscedastic, the method of weighted least squares can be applied. According to this method, the weighted sum of squares $\mathbf{e'D^{-1}e}$ is minimised. As in Section 3.1, the matrix $\mathbf{D}$ denotes the variance–covariance matrix of the calibrators. It is well known that the solution to the weighed least squares problem, in other words the estimates that minimise the weighted sum of squares, can be found by methods for ordinary least squares. If $\mathbf{C}$ is such that $\mathbf{D} = \mathbf{CC'}$ and the model $\mathbf{y} = f(\boldsymbol{\beta}) + \mathbf{e}$ is multiplied from the left by $\mathbf{C^{-1}}$, then the variance of $\mathbf{C^{-1}y}$ is an identity matrix, and the solution to the transformed model, $\mathbf{C^{-1}y} = \mathbf{C^{-1}}f(\boldsymbol{\beta}) + \mathbf{C^{-1}}(\mathbf{e})$, is the solution to the weighted least squares problem (Seber & Wild, 1989, p. 28).

In practice, the variance–covariance matrix $\mathbf{D}$ is not known. It could perhaps be assumed that the variance follows the power-of-the-mean model (2.5), or the log-linear model (2.6). In these models, the variance parameters, *i.e.* θ and ϕ, need to be estimated. In the power-of-the-mean model, the mean, *i.e.* E(y), also needs to be estimated.

The calibration curve is usually fitted to a small number of calibrators, often measured in duplicate. Because calibration datasets are small, estimation of variance parameters based on single calibration datasets cannot be recommended. If single calibration datasets are to be used for estimation of variance parameters, the datasets should be large, and checks should be made for erroneous measurements and outliers that could influence the estimates. When the single calibration datasets are small, the variance parameters can be estimated based on a collection of datasets. This method requires stable variance parameters that do not change from one calibration dataset to another, an assumption that is not completely realistic. On the

other hand, minor changes in the values of the variance parameters can have small effects on the inverse predictions of the sample concentrations. Zeng & Davidian (1997a) provide a statistical test for the hypothesis of equal variance parameters in calibrator datasets.

Regression analysis is the simplest method for estimation of the variance parameters θ and ϕ in (2.5) or (2.6). It requires replicate measurements of the calibrators. For each calibrator concentration, the average response is calculated and the intra–assay variance var(y) is estimated. In (2.5), regression of log var(y) on log-transformed averages gives estimates of the intercept, *i.e.* log ϕ, and the slope θ. This popular method was suggested by Rodbard *et al.* (1976). Similarly, in (2.6), regression of log var(y) on calibrator concentrations x gives estimates of the intercept log ϕ and the slope θ. Weightings can be used in the regression to account for heterogeneity in the variance of log var(y) (Rodbard *et al.*, 1976; Raab, 1981).

Raab (1981) proposed estimation of variance parameters by a modified maximum likelihood method, since the estimators based on regression analysis (Rodbard *et al.*, 1976) and the maximum likelihood estimators (Finney & Phillips, 1974) proved to be biased, especially in datasets with varying number of replicates. This iterative modified maximum likelihood method is more efficient than the regression method, but also more complicated. Sadler & Smith (1986) proposed that the modified maximum likelihood method be simplified by estimating expected responses by averages, making the method computationally faster. Sadler (2002) described an improved computer programme that makes use of this technique. These maximum likelihood methods are based on the normal distribution.

More generally, the variance can be written

$$\text{var}(y) = \phi\, g(x, \boldsymbol{\beta}, \boldsymbol{\theta}), \tag{3.1}$$

where the form of the function g is usually assumed to be known (Davidian, 1990). In (3.1), $\boldsymbol{\beta}$ denotes the curve parameters, and ϕ and $\boldsymbol{\theta}$ the variance parameters. When the variance-covariance matrix $\mathbf{D}$, or the estimate of $\mathbf{D}$, depends on $\boldsymbol{\beta}$, the weighted least squares method for estimation of $\boldsymbol{\beta}$ can be applied iteratively. A preliminary estimate of $\boldsymbol{\beta}$ makes it possible to estimate the variance parameters, producing a first estimate of $\mathbf{D}$. Using this first estimate of $\mathbf{D}$, an estimate of $\boldsymbol{\beta}$ can be obtained by weighted least squares, and the process is repeated.

The estimation of the variance parameters ϕ and $\boldsymbol{\theta}$ in (3.1) can be based on sample variances or residuals. For example, the regression method proposed by Rodbard *et al.* (1976) is based on sample variances. Such

24

methods require replicates but do not need any specification of the mean function. Goos, Tack & Vandebroek (2001) discussed designs for estimation of variance parameters, based on sample variances. Methods based on residuals are more efficient than methods based on sample variances (Davidian & Carroll, 1987), especially when the numbers of replicates are small, as is often the case in datasets from analytical procedures. Pseudolikelihood maximisation is a common method for variance parameter estimation that is based on residuals. According to this method, the variance parameters are estimated by maximum likelihood, conditioned on the current estimates of the curve parameters. Davidian & Giltinan (1993) and Giltinan & Davidian (1994) discussed this and other methods for variance parameter estimation in datasets including many assay runs. Davidian (1990) considered estimation of variance parameters when the data are possibly nonnormal, with unequal numbers of replicates.

Obviously, there are many possible variance functions, and many methods for estimating their parameters, producing different estimates. Fortunately, the choice of method usually has negligible effect on the estimates of the curve parameters. As explained by Carroll (2003) 'as the level of simply fitting the mean structure, *it does not matter what method is used to account for the variance structure*.' In order to model the mean structure correctly, however, it is necessary to account for the variance heterogeneity by some method.

Estimated variance parameters can be used for assessing the precision of the calibration. Giltinan & Davidian (1994) suggested that estimates of variance parameters be used for making intra-assay precision profiles. Similarly, estimates of variance parameters can be used for construction of sample concentration confidence intervals. There are four basic approaches for making such confidence sets:

    i.   Inversion of asymptotic prediction intervals (Fieller, 1954)
    ii.   Symmetric Wald intervals, based on the asymptotic variance in the predicted sample concentration
    iii.   Likelihood methods (Brown & Sundberg, 1987; Giltinan & Davidian, 1994; Bellio 2003)
    iv.   Bootstrap methods (Zeng & Davidian, 1997b; Jones & Rocke, 1999; Benton, Krishnamoorthy & Mathew, 2003).

Approach i) gives unsymmetrical intervals and may for this reason be more appropriate than approach ii) when the nonlinear function has an asymptote (Schwenke & Milliken, 1991). Belanger, Davidian & Giltinan

(1996) studied approaches i) an ii) and found that the confidence intervals depend critically on the quality with which the variance parameters are estimated. They concluded that 'the common practice of setting variance parameters to fixed values without adequate investigation may lead to erroneous calibration inference.' Zeng and Davidian (1997c) considered construction of confidence intervals based on information from many assay runs.

These methods for the construction of precision profiles for intra-assay variation and for the construction of confidence intervals generally presume that the variation in the samples is the same as the variation in the calibrators. In practice, this assumption is not always accurate. The calibrators are as a rule manufactured, and for this reason probably have smaller variance than genuine clinical samples. Furthermore, it is well known that predicted sample concentrations can differ considerably between *e.g.* laboratories, instruments and batches of reagents even if each assay is calibrated. This inter-assay variation is often larger than the intra-assay variation. For this reason, confidence intervals based solely on intra-assay precision, perhaps expressed by estimated variance parameters, are likely to be too narrow.

An investigation of precision should include many variance components, for example several instruments and batches of reagents, and the assays should be performed under varying conditions, for example under different temperatures. All relevant sources of errors should be considered when making a statement about the precision in a predicted sample concentration. The results from a well performed precision study, including a number of samples covering the range of measurement, could be used for assessing the precision in subsequent determinations of concentrations in clinical samples.

3.2.2 Inverse regression

By the method of inverse regression, the curve is fitted to concentrations, as opposed to responses. The sum of squared deviations between the points and the curve is minimised, with deviations measured on the concentration axis. This method contrasts with the classical method of calibration, which is the ordinary method of least squares, minimising the sum of squares measured on the response axis. Krutchkoff (1967) initiated a discussion by reporting results from a simulation study indicating a smaller mean square error in predicted sample concentrations by inverse regression than by classical regression. Berkson (1969) objected that the inverse predictor of sample concentrations is consistent in classical regression, but inconsistent in inverse regression. Tellinghuisen (2000) pointed out that the inverse

predictor is biased in finite datasets, even in classical regression, and for small datasets (approx. 5 observations) the bias in classical regression is comparable to the bias in inverse regression. Krutchkoff (1969) showed that classical regression can give a smaller mean square error than inverse calibration outside the calibration range. It is now well accepted that inverse regression can produce a smaller mean square error than classical regression in a range around the average concentration (Chow & Liu, 1995, p. 39). Tellinghuisen (2000) observed that the range of concentrations over which inverse regression is more efficient than classical regression is greater for small data sets than for large. However, it should be noted that the differences between classical and inverse regression are very small unless the variation is large. Furthermore, the relative performance of one calibration estimator over the other is dependent on the distribution (Shalabh & Toutenburg, 2006). The research on inverse regression for calibration has been reviewed by Osborne (1991).

3.2.3 Other criteria for calibration

Calibration in terms of the method of weighted least squares was discussed above. Calibration can also be carried out by the method of maximum likelihood. When the responses are normally distributed, the method of least squares and the method of maximum likelihood give the same curve parameter estimates.

In assays with lognormal distributed measurements, it is often assumed that the response vector $\mathbf{y}$ follows a multiplicative model $\mathbf{y} = f(\boldsymbol{\beta})\mathbf{e}$, where $\mathbf{e}$ is a vector of random errors that belong to the lognormal distribution (2.7) with $\mu = 1$. In this case, the transform-both-sides method, introduced by Carroll & Ruppert (1984), can be used for calibration. If the logarithmic transformation is applied on the multiplicative model, the result is the additive model $\log \mathbf{y} = \log f(\boldsymbol{\beta}) + \log \mathbf{e}$, with random errors $\log \mathbf{e}$ that are normally distributed with expected value 0. This model is easily fitted by standard software for nonlinear regression. The log transformation is a member of the family of power transformations suggested by Box & Cox (1964),

$$h(y, \lambda) = \begin{cases} (y^{\lambda} - 1)/\lambda, & \lambda \neq 0 \\ \log y, & \lambda = 0 \end{cases}. \tag{3.2}$$

It is possible to estimate λ in (3.2), and transform the responses and the calibration model correspondingly. Bonate (2006) provides an introduction to the transform-both-sides method.

When the data are gamma distributed, the calibration curve f can easily be fitted using software for generalised linear models (McCullagh & Nelder, 1989), provided that f, or a monotonic differentiable function of f, is linear in the parameters. Through this method, the likelihood is maximised. For example, if the expected response $E(y)$ follows the Michaelis–Menten model (2.1), a generalised linear model with reciprocal link can be fitted to inverse calibrator concentrations, because then

$$\frac{1}{E(y)} = \frac{1}{\beta_2} + \frac{\beta_1}{\beta_2} x^{-1},$$

which is linear in the parameters $1/\beta_2$ and β_1/β_2.

Rocke and Lorenzato (1995) proposed fitting a calibration model that can be written

$$y = f(x) \exp(e_1) + e_2, \tag{3.3}$$

where e_1 is $N(0, \sigma_1^2)$ and e_2 is $N(0, \sigma_2^2)$. This model include two variation components: e_1 giving approximately constant coefficient of variation in response for high levels of concentration, and e_2 giving approximately constant standard deviation in response for low levels of concentrations. Because this variance pattern is often observed in immunoassay, model (3.3) may have many applications.

If there are random errors not only in response, but also in concentration, methods for errors-in-variables (*e.g.* Carroll, Ruppert & Stefanski, 1995) can be used for calibration. Concentration is a 'controlled observation' if its values are fixed in advance, while the unknown true values varies randomly around the fixed values. When the regressor is a controlled observation, the usual least-squares estimators of the parameters are unbiased in the linear case (Berkson, 1950), but biased in the nonlinear case (Cheng & Van Ness, 1999, p. 45).

3.3 Inference on the Coefficient of Variation

Observations from immunoassay are often distributed with a standard deviation that is approximately proportional to the level of the observations. The coefficient of variation is useful as a measure of the random variability. It is used for stating acceptance criteria for precision (DeSilva *et al.*, 2003), and is often reported in directions for use and in precision studies (e.g. Brunnée *et al.*, 1996). It is sometimes claimed that a method has

homogeneous precision, as measured by the coefficient of variation, in all parts of the range of measurement, while at other times it is reported that the precision is different in different parts of the range of measurement. When analysing data from immunoassay, coefficients of variation are often calculated and compared. This section reviews methods for making inference on the coefficient of variation.

In a sample with observations y_i, $i = 1, 2, \ldots, n$, of a random variable Y, the coefficient of variation is defined as $c = s/m$, where

$$m = \frac{1}{n}\sum_{i=1}^{n} y_i , \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - m)^2} . \qquad (3.4)$$

The sample coefficient of variation c is usually regarded as an estimator of the population coefficient of variation γ, defined as

$$\gamma = \frac{\sqrt{\mathrm{var}(Y)}}{\mathrm{E}(Y)} . \qquad (3.5)$$

In immunoassay, concentrations cannot be negative. Nevertheless, they can be approximately normally distributed, and the normal distribution can, and often is, used as a model. The modelling of nonnegative measurements by the normal distribution requires that the expected value be larger than 3 standard deviations. Otherwise, negative observations are likely to occur according to the model. Provided that the nonnegative data follow an approximate normal distribution, the population coefficient of variation γ is, as a consequence, smaller than $1/3$. When the sample coefficient of variation is larger than $1/3$ in replicated observations of a positive variable, nonnormality is suggested and the lognormal distribution (2.7) or the gamma distribution (2.8) might describe the data better.

3.3.1 Point estimators

Because the density of the average m is positive in a neighbourhood of 0, the expected value of the sample coefficient of variation c does not exist. When analysing variables that can only take positive values this is seldom a problem, since the probability of an average close to 0 is usually negligible.

As discussed in Section 2.3, if the observations are lognormal distributed, the variance calculated on log values approximately equals the squared coefficient of variation. The standard deviation, calculated on log values, is

for this reason a common estimator of the coefficient of variation when the data are lognormal.

When the data are gamma distributed, inference on the coefficient of variation amounts to inference on the shape parameter κ in the probability density function (2.8). This parameter can be estimated by maximum likelihood (Cohen & Whitten, 1982; Bowman & Shenton, 1983). A numerical method has to be applied, because there is no explicit solution. Standard statistical software for fitting generalised linear models (McCullagh & Nelder, 1989) can be used for simultaneous estimation of the expected value and the shape parameter by fitting a model including only an intercept. The sample mean is the maximum likelihood estimator of the expected value $\kappa\lambda$. Yanagimoto (1988) recommended estimation of κ by the conditional maximum likelihood estimator given the sample mean. The conditional maximum likelihood estimator of κ is consistent and less biased than the unconditional maximum likelihood estimator, but it requires more difficult calculations.

3.3.2 Single sample tests and confidence intervals

The test statistic

$$t = \frac{m}{s/\sqrt{n}} = \frac{\sqrt{n}}{c},$$

for the hypothesis that the expected value of a normally distributed random variable equals 0 is t distributed with $n - 1$ degrees of freedom, provided that the hypothesis is true. Generally t follows a noncentral t distribution with $n - 1$ degrees of freedom and noncentrality parameter $\tau = n^{1/2}/\gamma$. Owen (1968) discussed this and other applications of the noncentral t distribution. A confidence set for τ can be constructed from the acceptance region of a test of the hypothesis about τ. Thus, if $\Pr(t < n^{1/2}/c \mid \tau = \tau_1) = \alpha/2$ and $\Pr(t > n^{1/2}/c \mid \tau = \tau_2) = \alpha/2$ then $[\tau_2, \tau_1]$ is a $100(1 - \alpha)$ % confidence interval for τ. An exact finite confidence interval for γ is easily obtained from the confidence interval for τ provided that the latter does not include 0. The exact finite confidence interval is

$$\left[\frac{\sqrt{n}}{\tau_1} \quad , \quad \frac{\sqrt{n}}{\tau_2} \right]. \tag{3.6}$$

In practice, the percentiles of the noncentral t distribution might not be available, because computer programmes for statistics do not always include

the noncentral t distribution. The open source software R, version 7.2.2, produces incorrect results for noncentrality parameters smaller than 37.62 (R Development Core Team, 2008, p. 1334). Fortunately, there are several ways to calculate approximate confidence intervals. Miller (1991a) showed that c is asymptotically normally distributed with expected value γ and variance

$$\frac{\gamma^2(1+2\gamma^2)}{2(n-1)} \quad . \tag{3.7}$$

Based on (3.7), Miller & Feltz (1997) proposed the approximate $100(1 - \alpha)$ % confidence interval

$$\left[c - z\sqrt{\frac{c^2(1+2c^2)}{2(n-1)}} \quad , \quad c + z\sqrt{\frac{c^2(1+2c^2)}{2(n-1)}} \right], \tag{3.8}$$

where z is the $100(1 - \alpha/2)$:th percentile of the standard normal distribution. This confidence interval is symmetrical around c. A similar single sample test of the null hypothesis $\gamma = \gamma_0$ was proposed by Rao & Bhatta (1989), who also made an Edgeworth expansion of the null distribution of the sample coefficient of variation. An asymmetrical interval, more likely than (3.8) to perform well also for smaller sample sizes, is obtained if only the second γ^2 in (3.7) is estimated by c^2. Thus

$$\frac{c - \gamma}{\gamma\sqrt{(1+2c^2)/(2(n-1))}}$$

is approximately distributed as a standard normal distribution. The corresponding $100(1 - \alpha)$ % confidence interval for γ is

$$\left[\frac{c}{1 + z\sqrt{(1+2c^2)(2(n-1))}} \quad , \quad \frac{c}{1 - z\sqrt{(1+2c^2)(2(n-1))}} \right], \tag{3.9}$$

which, according to Reh & Scheffler (1996), was suggested by Graf *et al.* (1987). Hald (1952) proposed another approximate confidence interval based on asymptotic normality.

McKay (1932) showed that if $\gamma < 1/3$ and $\theta = (n - 1)/n$, then

$$\frac{(n-1)\ (1+\gamma^2)\ c^2}{(1+\theta\ c^2)\ \gamma^2} \tag{3.10}$$

is approximately χ^2 distributed with $n-1$ degrees of freedom. As explained in the introduction to Section 3.3, in applications the condition $\gamma < 1/3$ is usually reasonable, since it makes negative observations unlikely. Let u_1 denote the $100(1-\alpha/2)$:th percentile, and u_2 denote the $100\alpha/2$:th percentile of a χ^2 distribution with $n-1$ degrees of freedom. Because (3.10) is an approximate pivotal quantity, it can be used for calculating an approximate confidence interval,

$$\left[\frac{c}{\sqrt{u_1/(n-1)+c^2(u_1/n-1)}} \quad , \quad \frac{c}{\sqrt{u_2/(n-1)+c^2(u_2/n-1)}}\right]. \tag{3.11}$$

Of course, as pointed out by David (1949), (3.10) could also be expected to be approximately χ^2 distributed with $n-1$ degrees of freedom if $\theta = 1$. Vangel (1996) showed that

$$\theta = \frac{n-1}{n}\left(\frac{2}{u_2}+1\right)$$

is optimal for calculating the $100\alpha/2$:th percentile of the sample coefficient of variation. The approximate confidence interval based on this result is

$$\left[\frac{c}{\sqrt{u_1/(n-1)+c^2((u_1+2)/n-1)}}, \quad \frac{c}{\sqrt{u_2/(n-1)+c^2((u_2+2)/n-1)}}\right], \tag{3.12}$$

with u_1 and u_2 defined as in (3.11).

Assuming that the data are lognormal, the coefficient of variation on the original scale can be estimated by the standard deviation on the log scale. The usual confidence interval for a standard deviation calculated on log values is thus an approximate confidence interval for the coefficient of variation in the original values. This confidence interval can be written

$$\left[\sqrt{\frac{(n-1)s_L^2}{u_1}} \quad , \quad \sqrt{\frac{(n-1)s_L^2}{u_2}}\right], \tag{3.13}$$

where s_L denotes the standard deviation calculated on log values.

The following example includes the 'naive' interval

$$\left[\sqrt{\frac{(n-1)c^2}{u_1}} \quad , \quad \sqrt{\frac{(n-1)c^2}{u_2}} \right]. \tag{3.14}$$

This interval is obtained if the coefficient of variation is treated as if it were a proper standard deviation, *i.e.* if the limits of the usual confidence interval for σ^2 are divided by the average m. Vangel (1996) compared analytically the errors in this naive approximation with the error in McKay's approximation and concluded that the naive approximation is 'substantially less accurate'.

In the example below approximate confidence intervals are calculated on tensile strength data given by Vangel (1996). The data set, presented in Table 3.1, consists of five measurements.

Table 3.1. Measurements of tensile strength provided by Vangel (1996)

Measurement	Tensile strength (1000 psi)
1	326
2	302
3	307
4	299
5	329
Mean (1000 psi)	312.6
Coefficient of variation	0.0446

The calculated confidence intervals are given in Table 3.2. The exact confidence interval was obtained using the function `tnonct` in SAS 9.1 (SAS Institute Inc., Cary, NC, USA).

Table 3.2. 95% confidence intervals for the coefficient of variation in the tensile strength data

Method (eq. no.)	Confidence interval
Exact (3.6)	[0.0267 , 0.1287]
Miller & Feltz (3.8)	[0.0136 , 0.0756]
Graf *et al.* (3.9)	[0.0263 , 0.1459]
McKay (3.11)	[0.0267 , 0.1291]
Vangel , (3.12)	[0.0267 , 0.1287]
Log (3.13)	[0.0266 , 0.1274]
Naive (3.14)	[0.0267 , 0.1281]

In this dataset the estimated coefficient of variation was small (4.46%). Most of the approximate confidence intervals were similar to the exact confidence interval. However, the symmetrical Miller & Feltz confidence

interval did not perform well (Table 3.2). The confidence interval suggested by Graf *et al.* (1987), which is based on the same normal approximation, was more adequate. The McKay approximation proved to be very accurate, and the modification due to Vangel (1996) was successful. The method of logarithmic transformation worked adequately in this example but not as good as McKay's approximation. The good performance of McKay's approximation (3.10) is not surprising. Fieller (1932), Pearson (1932), Iglewicz & Myers (1970) and Umphrey (1983) have confirmed that (3.10) is approximately χ^2 distributed. The naive method also performed well, but the interval was somewhat too small since the method ignores the variation in the estimate of the average.

The robustness of inferential methods for the coefficient of variation based on McKay's approximation has been little investigated. Bonett & Seier (2006) noted that the confidence interval suggested by Vangel (1996) performed poorly with nonnormal distributions. They proposed, as an alternative, an approximate confidence interval for the coefficient of dispersion (*i.e.* the mean absolute deviation from the median, divided by the median).

The arithmetic and geometric sample means are the complete and sufficient statistics for the parameters of the gamma distribution (2.8). Confidence intervals for the shape parameter κ in (2.8) can be based on the ratio between these means (Engelhardt & Bain, 1978). Because this ratio does not depend on the scale parameter λ in (2.8), it is easy to simulate critical values for a test of the hypothesis that κ equals some specific value. Likelihood methods for constructing confidence intervals and calculating approximate tail probabilities in inference on the shape parameter were investigated by Wong (1992) and Wong & Wu (1998). Wong & Wu (2002) proposed an approximate method for calculating confidence intervals for the population coefficient of variation γ based on the modified signed log likelihood ratio statistic defined by Barndorff-Nielsen (1986, 1991). This method was reported to give accurate results even in the case of small sample sizes, and can be used for gamma as well as normally distributed data. Simple large-sample methods for construction of tests and confidence intervals for the inverse sample coefficient of variation $1/c$ in gamma and Weibull distributed data, based on asymptotic normality, were proposed by Sharma & Krishna (1994).

Likelihood ratio based confidence intervals for the shape parameter κ in the gamma distribution (2.8) can easily be obtained using software for fitting generalised linear models (McCullagh & Nelder, 1989). When there is only one sample, a model including only an intercept is appropriate. In this

34

approach, approximate confidence intervals can either be constructed from the estimated covariance matrix at the final iteration, or, more advanced, from investigation of the profile likelihood (e.g. Pawitan, 2001). In either case, the inference is based on asymptotic distributions, and may for this reason perform well in large datasets. When the observations are gamma distributed, the generalised linear model approach is convenient for estimating a common coefficient of variation, *i.e.* a population coefficient of variation that many samples have in common.

Bayesian theory provides methods for constructing posterior intervals (*i.e.* probability intervals) for the coefficient of variation. Pang *et al.* (2005) explained how this can be done when the observations are lognormal, gamma or Weibull distributed. They suggested the use of non–informative uniform prior distributions and the altering conditional sampling known as Gibbs sampler. This method is suitable when the distribution of the observations is known, but the probability density of the sample coefficient of variation is difficult to derive.

Rao & Bhatt (1995) described jackknife and bootstrap single sample tests for the coefficient of variation. The jackknife test statistic is asymptotically normal. By the bootstrap technique, critical values of a test statistic are determined by resampling from a parametric distribution or from the sample itself. More research is needed on the performances of jackknife and bootstrap methods for the coefficient of variation when the sample is small.

3.3.3 Tests for equality of coefficients of variation

When studying precision in immunoassay, samples with varying levels of concentration are measured in replicate, and their coefficients of variation are calculated. The coefficients of variation are often compared, usually informally, without performing any statistical tests. However, many methods have been proposed for making appropriate comparisons.

If it is possible to assume that the observations are lognormal distributed, variations could, with advantage, be compared on the log scale. After logarithmic transformation, the data are normally distributed, and the standard deviations equal the original coefficients of variation approximately (see Section 2.3). Standard methods for testing equality of variances can be applied. This approach can usually be recommended, because of ease and adequacy.

When it is believed that the distribution is symmetrical, and when other statistical analyses performed on the data, such as regression and analysis of variance, are based on the normal distribution, there could be an interest in other approaches. However, when the coefficients of variation are small, the

lognormal distribution approximates normal distributions well (Figure 2.3), and the differences between analyses based on the lognormal distribution and analyses based on the normal distribution are small.

Several authors have explored the likelihood ratio test for normally distributed data. Miller & Karson (1977) and Bhoj & Ahsanullah (1993) considered the special case of equal sample sizes. Lohrding (1975), Bennett (1977) and Doornbos & Dijkstra (1983) treated the general case of unequal sample sizes. The likelihood ratio test is computationally inconvenient, because there are usually no explicit expressions for the maximum likelihood estimates under the null hypothesis of equal coefficients of variation. Numerical procedures for the likelihood ratio test have been proposed by Gupta & Ma (1996) and Nairy & Rao (2003). When only two coefficients of variations are to be compared the likelihood ratio test can be written explicitly; because in this case exact expressions are available for the maximum likelihood estimates under the null hypothesis (Gerig & Sen, 1980). The likelihood ratio test is known to be too liberal for small sample sizes (Doornbos & Dijkstra, 1983; Fung & Tsang, 1998; Nairy & Rao, 2003). In order to improve the performance of the likelihood ratio test, Verrill & Johnson (2007) provided an internet based programme that simulates critical values. Gupta & Ma (1996) derived a score test, based on maximum likelihood estimates, and a Wald test for unequal sample sizes, build on the Wald test for equal sample sizes proposed by Rao & Vidya (1992). A similar test, for equal sample sizes, was proposed by Bhoj & Ahsanullah (1993). Doornbos & Dijkstra (1983), Singh (1993), Sharma & Krishna (1994) and Nairy & Rao (2003) proposed tests based on the distribution and the moments of the inverse sample coefficient of variation $1/c$.

Feltz and Miller (Miller, 1991a; Feltz & Miller, 1996; Miller & Feltz, 1997) have contributed a test, based on asymptotic normality, for equality of k coefficients of variation. Let c_i be the sample coefficient of variation in population i, calculated on n_i observations, $i = 1, 2, \ldots, k$, and let $c = \Sigma_i c_i(n_i - 1)/\Sigma_i (n_i - 1)$, and $N = \Sigma_i n_i$. The test statistic,

$$\frac{\sum_{i=1}^{k}(n_i - 1)\ (c_i - c)^2}{c^2(1/2 + c^2)}, \tag{3.15}$$

is approximately χ^2 distributed with $N - k$ degrees of freedom.

Bennett (1976) utilised McKay's approximation (3.10) for the coefficient of variation. Because (3.10) is approximately χ^2 distributed, and

consequently approximately gamma distributed, Bennett (1976) suggested that Pitman's test (Pitman, 1939) for equality of scale parameters in gamma distributions be applied. Shafer & Sullivan (1986) noted that Bennett, when deriving the test statistic for the coefficient of variation, mistakenly used a sample variance with divisor $n - 1$ whereas McKay (1932) used a sample variance with divisor n. For this reason they modified Bennett's test correspondingly. The modified Bennett's test statistic for equality of coefficients of variation is

$$F = (N-k)\log\left(\frac{\sum_{i=1}^{k}(n_i-1)\ u_i}{N-k}\right) - \sum_{i=1}^{k}(n_i-1)\log u_i, \qquad (3.16)$$

where $u_i = c_i^2/(1 + c_i^2(n_i - 1)/n_i)$. The test statistic (3.16) is approximately χ^2 distributed with $N - k$ degrees of freedom.

Wilson & Payton (2002) utilised McKay's approximation in a similar way. They noted that because McKay's approximation is approximately gamma distributed, it can be used for modelling the coefficient of variation by application of iteratively weighted least squares in the framework of generalised linear models (McCullagh & Nelder, 1989). This makes it possible to investigate hypotheses about the coefficient of variation by use of asymptotically valid likelihood ratio tests. The method proposed by Wilson & Payton (2002) is useful, since the coefficient of variation can be modelled by categorical factors and regression terms as in ordinary linear models, but the statistical tests may require large datasets.

Tests for equality of coefficients of variation in gamma distributed data can be based on the ratio between the arithmetic and the geometric sample means (Keating, Glaser & Ketchum, 1990; Yanagimoto & Yamamoto, 1991). Alternatively, the generalised minimum chi-square procedure proposed by Tripathi, Gupta & Pair (1993) can be applied, with attention to the correction in Bhattacharya (2002).

Miller (1991b) suggested a nonparametric test for equality of coefficients of variation. This nonparametric test was recommended for nonnormal distributions by Fung & Tsang (1998).

4 Results

The main ideas and results of the thesis are summarised in this section with references to Papers I-IV, which are included as appendices. Paper I, which considers the problem of choosing calibrator concentrations, is discussed in Section 4.1. Paper II, on inverse regression for immunoassay, is reported in Section 4.2. Papers III and IV, which deal with inference on the coefficient of variation and with the distribution of McKay's approximation, are considered together in Section 4.3.

4.1 Calibration Design

Paper I introduces a new method for determining optimal designs. It proposes that the existing methods for finding locally optimal designs (François, Govaerts & Boulanger, 2004) be extended by allowing for randomness in the curve parameters. As discussed, in Section 3.1, the Bayesian approach (Chaloner & Verdinelli, 1995) requires specification of the full multivariate distribution of the curve parameters. The new method, described in Paper I, only requires an assumption about the expected value $\boldsymbol{\beta}_0$ of the curve parameters and their variance-covariance matrix $\boldsymbol{\Sigma}$. The matrix $\boldsymbol{\Sigma}$ includes inter-assay variances and covariances for the distribution of the parameter values. The parameter values vary, for example as a consequence of inter-assay changes in temperature, variation in assay execution or differences between assays in batches of reagents. The variance-covariance matrix $\boldsymbol{\Sigma}$ should not be confused with the variance-covariance matrix $\mathbf{V}$, which includes variances and covariances in the parameter estimators, caused by intra-assay variation.

Paper I defines the design criterion as a function h of the variance $\mathrm{var}(x)$ and the expected value $\mathrm{E}(x)$ in the predicted concentrations x, integrated over the range of measurement. This function can be the variance or the

mean square error. For immunoassay, it is suggested that h be the coefficient of variation and that the area under the precision profile be minimised. This makes sense when the magnitude of the relative errors is more interesting than the magnitude of the absolute errors.

The design criterion can be written

$$\int h_\xi \pi_\xi d\xi \,,$$

where ξ denotes the true sample concentrations, and π is a probability density function of ξ. By choosing the probability density function π, the function h can be integrated over the range of measurement on the original scale or on a log scale. It is also possible to give larger weightings to some parts of the range of measurement than to others. For example, large weightings can be assigned to concentrations close to clinical cut-off values, and small weightings can be assigned to clinically less interesting parts of the range of measurement.

Because in nonlinear regression the variance and the expected values in the inverse predictions are dependent on the parameter values, which are unknown, Paper I suggests that the design criterion be minimised as calculated by the equations for unconditional expected values and variances:

$$\mathrm{E}(x) = \mathrm{E}(\mathrm{E}(x|\beta)),$$

$$\mathrm{var}(x) = \mathrm{E}(\mathrm{var}(x|\beta)) + \mathrm{var}(\mathrm{E}(x|\beta)).$$

These expressions can be calculated approximately by series expansion about $\beta = \beta_0$, and by making use of the variance-covariance matrix Σ. In contrast, locally optimal designs are based on conditional expected values and variances. For this reason, the method proposed in Paper I can be regarded as an extension of the method for finding locally optimal designs.

4.2   Calibration Criteria

Inverse regression was discussed in Section 3.2.2. Paper II proposes weighted inverse regression for calibration in immunoassay with proportional errors. According to this method the calibration curve is fitted inversely, and with reciprocals of squared calibrator concentrations as weightings. The basis for this suggestion is the notion that the random variation in the dispensed volumes of the calibrators and the clinical samples could be the main source of error in advanced diagnostic measuring

instruments. Varying volumes give varying numbers of molecules included in the sample. This is illustrated in Figure 4.1, where the varying large circles indicate varying volumes, and the smaller filled circles symbolise molecules. The number of molecules included in the sample varies much more when the concentration is high than when the concentration is low. Paper II shows that the standard deviation in the number of molecules is proportional to the concentration, resulting in a constant coefficient of variation. The main contributions of Paper II are this plausible reason for the approximate homogeneity in coefficients of variation in immunoassay, and its consequences for curve fitting.
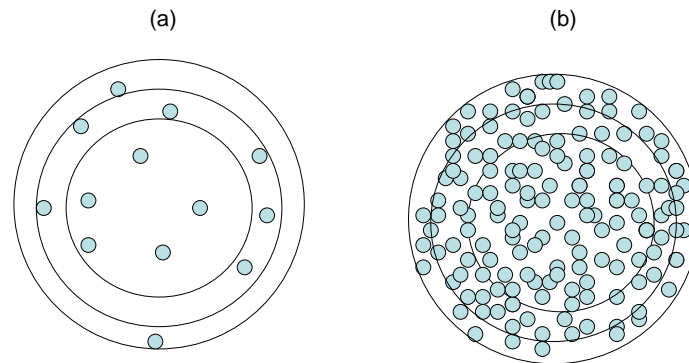

Figure 4.1. Varying sample concentrations. a) Low concentration b) High concentration.

Provided there is a constant coefficient of variation on the concentration axis, rather than on the response axis, weighted inverse regression is suggested, because it minimises relative squared residuals in concentration. This criterion also agrees with the common use of the coefficient of variation for measuring variability in immunoassay. When the concentrations are normally distributed, the maximum likelihood method is equivalent to the method of inverse regression (Paper II).

If the response is proportional to the number of molecules in the sample, a constant coefficient of variation in the number of molecules implies a constant coefficient of variation in response. Therefore, it is not surprising that often $\theta \approx 2$ in the power-of-the-mean model (2.5) for the variance. Paper II shows that if the response is not proportional to the concentration, but follows the four-parameter logistic function (2.2), then the standard deviation in response is approximately

$$\frac{\beta_4(y - \beta_1)(\beta_2 - y)\gamma}{(\beta_2 - \beta_1)}, \qquad\qquad (4.1)$$

where γ is the coefficient of variation in concentration and y denotes the response level.

As mentioned in Section 3.2.1, Carroll (2003) noted that calibration curves are only slightly influenced by the choice of method for taking heteroscedasticity into account. This is illustrated by the example in Paper II, which includes a comparison between three methods for calibration. Two of the methods, weighted classical regression and weighted inverse regression, account for the heteroscedasticity and give similar results. The third method is an ordinary least squares fit of the calibration curve, without use of weightings. This method differs considerably from the other methods with regard to predicted sample concentrations.

The main arguments for weighted inverse regression in immunoassay, compared with weighted classical regression, are that:

i. It has a theoretical basis (Paper II).
ii. It focuses on minimising errors in concentration rather than in response.
iii. It does not require estimation of variance parameters.

Argument iii) makes weighted inverse regression convenient, because, as discussed in Section 3.2.1, it may be difficult to decide upon a method for estimation and to determine how frequently the estimates should be updated.

4.3   Inference on the Coefficient of Variation

It is often appropriate to assume that immunoassay measurements are lognormal distributed, and this is convenient when investigating the variation, as noted in Section 3.3. However, in some applications, there could be reasons for believing that the observations are normally distributed. The theoretical argument given in Paper II to some extent supports the assumption of symmetrically distributed measurements in immunoassay. The random variation in the sample volumes could be the main source of variation in highly automated diagnostic systems, and the distribution of the sample volume might be normal. This thesis contributes to the research about inference on the coefficient of variation, conditioned on the normal distribution.

Paper III shows that, when the distribution is normal, the sample coefficient of variation c is an asymptotically unbiased estimator of the population coefficient of variation γ. It also shows, by Taylor series expansion, that if the population coefficient of variation γ is small, c can be anticipated to be close to $\gamma(1 - 1/(4(n - 1)))$, where n is the number of observations. In immunoassay, γ is usually smaller than 0.20, and for this case it is proposed that c divided by $1 - 1/(4(n - 1))$ be used as a bias adjusted estimator of γ.

In immunoassay datasets, the coefficient of variation is frequently calculated for many subsets of the data, for example for many samples. When these are similar, a method for summarising them into one estimate of a common coefficient of variation is required. Paper III proposes that the squared coefficients of variation be pooled by the degrees of freedom (*i.e.* the number of observations minus 1), and that the square root of this pooled average be used as an estimator of γ, after bias adjustment. This is a simple procedure that may have many applications.

Furthermore, when there are samples from many populations that have the same population coefficient of variation but possibly different expected values, a confidence interval for the common population coefficient of variation may be required. Tian (2005) proposed a method for calculating confidence intervals, based on resampling. Verrill & Johnson (2007) provided a method based on the likelihood ratio test with critical values simulated in an application on the internet. Paper III proposes that McKay's approximation be used for calculating confidence intervals for common coefficients of variation. Explicit equations that are easy to account for and that give the same result whenever applied, in contrast to resampling methods, are provided.

Paper III suggests that McKay's approximation (3.10) be employed straightforwardly for the comparison of two coefficients of variation c_1 and c_2, calculated on n_1 and n_2 observations, respectively. A test similar to the usual F-test for the comparison of two variances is proposed:

$$\frac{c_1^2/(1+c_1^2(n_1-1)/n_1)}{c_2^2/(1+c_2^2(n_2-1)/n_2)} \ . \tag{4.2}$$

The test statistic (4.2) is approximately F distributed with $n_1 - 1$ and $n_2 - 1$ degrees of freedom (Paper III). This approximate F-test performs well, according to the simulation study in Paper III, especially with regard to type I error, even when the sample sizes are very small.

If more than two coefficients of variation are to be compared, it is proposed that this be done by use of Hartley's F-test (Hartley, 1950). Then c_1 denotes the largest coefficient of variation and c_2 the smallest, and the test statistic (4.2) is compared with the critical values listed by Nelson (1987). Paper II includes an example of this method.

Unlike many other tests, the approximate F-test proposed in Paper III can be used when there are many estimates per population coefficient of variation. In other words, it can be used for comparisons of common coefficients of variation. When studying precision in immunoassay it is often interesting to compare coefficients of variation for different groups of samples. One group of samples could have one population coefficient of variation in common, and a second group of samples could have a second population coefficient of variation in common. With the methods in Paper III it is possible to statistically test the equality of the two common population coefficients of variation.

Feltz & Miller's test (3.15) and Bennett's test (3.16) perform well with regard to type I and type II error, and have been recommended for this reason (Feltz & Miller, 1996; Gupta & Ma, 1996; Nairy & Rao, 2003). In the simulation study in Paper III, the approximate F-test (4.2) was compared with the likelihood ratio test, Feltz & Miller's test (3.15) and Bennett's test (3.16). The approximate F-test produced accurate frequencies of falsely rejected null hypotheses (approx. 5%), even in cases of small sample sizes, and the powers were comparable to those of Feltz & Miller's test (3.15) and Bennett's test (3.16). As expected (cf. Section 3.3.3), the likelihood ratio test proved to be too liberal.

The assumption of the distribution is crucial when the coefficient of variation is large. The approximate F-test is appropriate for the comparison of two coefficients of variation, provided that the observations are normally distributed. If the observations are lognormal, an ordinary F-test should be applied after logarithmic transformation of the observations. This can be called a lognormal test. The importance of the assumption about the distribution is illustrated in Figures 4.2 *a-c*, which add results to the simulation study presented in Paper III. Three cases were investigated, with a common population coefficient of variation $\gamma_1 = \gamma_2 = \gamma$ equal to 5%, 15% and 30%, respectively. Normally distributed samples of n_1 and n_2 observations, respectively, were generated in MATLAB, version 6.5 (MathWorks, Natick, MA, USA), with $n_1 = n_2 = n$, $n = 2, 3, \ldots, 20$ (*i.e.* the sample sizes were always equal and varied from 2 to 20). Thus 19 simulations were made per case. In each simulation, 20,000 pairs of samples were generated. The results of the approximate F-test are presented in full in

Figures 1–3 in Paper III, and are included in Figures 4.2 *a-c* for comparison with the results of the lognormal test.
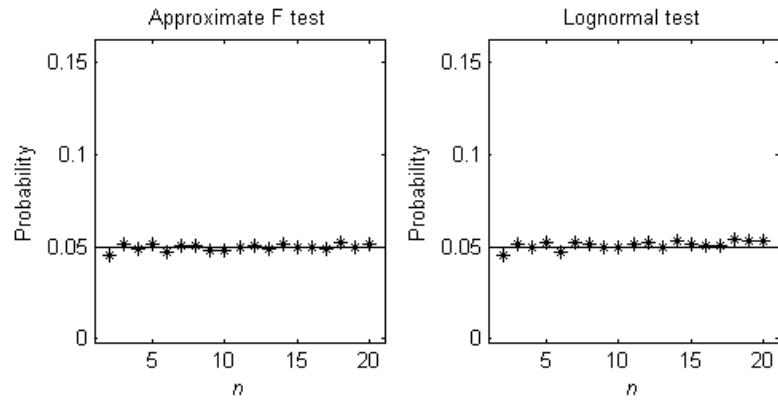

Figure 4.2a. Comparison of the approximate F-test (4.2) and the lognormal test on simulated normally distributed observations. Probability of type I error when $\gamma = 5\%$ and $n = n_1 = n_2$.

Figure 4.2b. Comparison of the approximate F-test (4.2) and the lognormal test on simulated normally distributed observations. Probability of type I error when $\gamma = 15\%$ and $n = n_1 = n_2$.
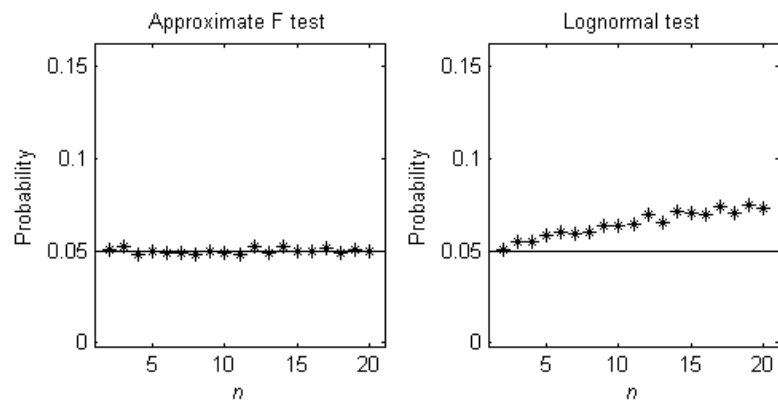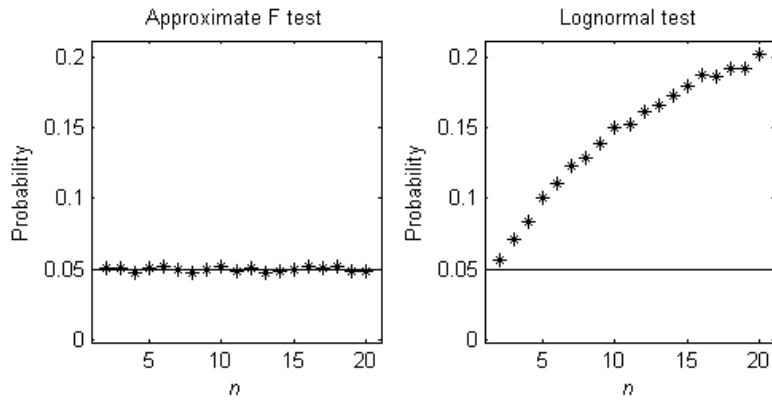
Figure 4.2c. Comparison of the approximate F-test (4.2) and the lognormal test in simulated normally distributed observations. Probability of type I error when $\gamma = 30\%$ and $n = n_1 = n_2$.

Evidently, the lognormal test and the approximate F-test perform similar with regard to type I error when the coefficient of variation is small (Figure 4.2 *a*), but not otherwise (Figures 4.2 *b* and *c*). The frequency of falsely rejected hypotheses of equal coefficients of variation reached 20% when the common population coefficient of variation was 0.30, and the sample size was large ($n = 20$). Note that the upper limit of the y-axis is 20% in Figure 4.2 *c*, but 15% in Figures 4.2 *a* and *b*.

It had been assumed that McKay's approximation is 'asymptotically exact' (Vangel, 1996), but Paper IV shows it is asymptotically normal with mean $n - 1$ and variance $2(n - 1)(1 + 2\gamma^2)/(1 + \gamma^2)^2$. However, as Paper IV discusses, because γ is small, the asymptotic variance does not differ greatly from $2(n - 1)$, and for this reason McKay's approximation should asymptotically perform sufficiently well in most applications. Paper IV also shows that McKay's approximation is type II noncentral beta distributed. Paper III utilises this result for providing exact expressions for the first and second moments of the approximation. It is noticed, for some small sample sizes, that the first and second moments approximately equal the first and second moments of a χ^2 distributed random variable with $n - 1$ degrees of freedom.

46

5   Conclusions

This concluding section lists the main contributions of this work (Section 5.1), contains some final remarks (Section 5.2), and discusses potential future work (Section 5.3).

5.1   Main Contributions

The main contributions of this thesis can be summarised in the following points:

i. A new approach for designing nonlinear calibration was developed. This new approach can be regarded as an extension of the methods for finding locally optimal designs. Through this extension, inter-assay variation in the curve parameters can be taken into account (Paper I).

ii. Theoretical arguments were presented for a homogeneous coefficient of variation in concentration and for using weighted inverse regression for calibration in immunoassay (Paper II).

iii. An approximate formula was provided for the standard deviation in response as a function of the level of the response, under assumptions of a four-parameter logistic function and a constant coefficient of variation in concentration (Paper II).

iv. An estimator of a common coefficient of variation and its asymptotic distribution were derived, and a bias-adjustment for small sample sizes was proposed (Paper III).

v. The use of McKay's approximation for constructing confidence intervals and testing hypotheses about common coefficients of variation in normally distributed data was demonstrated (Paper III).

vi. Exact expressions were given for the first two moments of McKay's approximation (Paper III).

vii.   The exact distribution of McKay's approximation for the coefficient of variation was determined (Paper IV).

viii. The asymptotic distribution of McKay's approximation for the coefficient of variation was derived (Paper IV).

5.2   Final Remarks

The new method for design of calibration, introduced in Paper I, allows for inter-assay variation in the curve parameters. In this regard, the method is an extension of the method for determining locally optimal designs. Application of the new method often reveals that the expected values of the curve parameters are more important than the covariance matrix. Designs that are optimal according to the new method are often locally optimal at the expected values of the curve parameters. In other words, the optimal design is sensitive to the assumption about the expected values. This is a problem, because the expected values may be hard to determine. Bayesian methods for finding optimal designs share this problem.

In immunoassay, the differences between classical regression and inverse regression are usually small. A weighted regression is needed when the standard deviation is increasing with the average of the measurements. The use of weighted inverse regression for calibration, as described in Paper II, is convenient because it does not involve estimation of variance parameters. However, knowledge about expected variances is important when assessing the residuals and the quality of the fit. Precision statements about inverse predictions of sample concentrations based on estimated variance parameters can be misleading, since they may not include all relevant variance components.

When making inference on coefficients of variation, it is practical to assume that the observations are lognormal distributed. Methods for lognormal distributed observations are applicable on approximately normally distributed positive measurements provided that the coefficient of variation is small, which is often the case in immunoassay. When the observations are positive and the coefficient of variation is small, the lognormal and the normal distributions are similar. The assumption about the distribution is critical when the coefficient of variation is large. Nonparametric methods, for example based on resampling, are useful when the distribution is unknown.

5.3 Future Research

Paper I proposed a method for determining optimal designs by linear approximations of unconditional expected values and variances. This idea was tested for the problem of determining design points for calibration in immunoassay. The idea might be applicable to other problems as well, as an alternative to Bayesian methods.

The model considered in Paper II was proposed under the assumption of negligible random variation in response. This assumption is often not realistic. The model needs to be extended to include substantial random errors in response. Methods for errors-in-variables, with concentration as a controlled variable, could be considered. Design issues for immunoassay calibration could be studied for errors-in-variables models.

Many statistical issues in immunoassay were not considered in this thesis, for example limits of detection and quantitation, conformity between immunoassay techniques, problems with interference and carry over, quality control in production, and estimation of sensitivity and specificity. Methods for these problems also need to be developed and tested.

The robustness of the methods proposed for inference on common coefficients of variation in Paper III remains to be examined. An extensive simulation study, comparing several tests for hypotheses about coefficients of variation under varying conditions, would be useful. Small sample properties of resampling methods for inference on the coefficient of variation could be investigated.

6 Sammanfattning

Statistiska metoder används bland annat för planering av vetenskapliga försök, modellering av samband och för att dra slutsatser av observationer. I den här avhandlingen utvecklas statistiska metoder för immunkemiska tekniker. I kemin används analytiska tekniker för att bestämma koncentrationer av kemiska föreningar. Immunkemiska tekniker är laborativa test, baserade på antikroppar, som mäter koncentrationerna av något ämne i patientprov. Flera patientprov brukar analyseras samtidigt i en så kallad körning

Eftersom immunkemiska tekniker är komplicerade test bestäms provens koncentrationer aldrig exakt, utan med fel som delvis kan betraktas som slumpmässiga. Det är vanligt att felens standardavvikelse är approximativt proportionell mot medelvärdet. Av denna anledning är variationskoefficienten, dvs. standardavvikelsen dividerad med medelvärdet, ett vanligt mått på spridning i immunkemi.

Avhandlingens utgångspunkt är att variationskoefficienten är ett relevant mått på spridning i immunkemiska metoder. Genom hela avhandlingen ligger fokus på relativa fel snarare än absoluta, eftersom relativa fel oftast är kliniskt mer intressanta. Avhandlingen syftar till att utveckla statistiska metoder för immunkemiska tekniker under denna förutsättning.

De immunkemiska testen ger en signal som är relaterad till provets koncentration. Signalen, som till exempel kan vara fluorescens, kallas vanligen respons. För att kunna översätta provets respons till koncentration är det nödvändigt att fastställa det aktuella sambandet mellan respons och koncentration. Detta kallas kalibrering.

Vid kalibrering används så kallade kalibratorer, vilka är prov med kända koncentrationer. Kalibratorerna är oftast inte naturliga, utan de tillverkas så att de får de rätta koncentrationerna. Normalt används ett fåtal kalibratorer med koncentrationer valda så att de täcker hela mätområdet. En kali-

breringskurva anpassas till kalibratorernas responser, normalt med minsta kvadratmetoden. Skattningen av ett patientprovs koncentration, via kalibreringskurvan, kallas invers prediktion. Den logisiska funktionen med fyra parametrar (2.2) används ofta för kalibrering av immunkemiska tekniker.

I följande avsnitt behandlas, med referenser till artiklarna I − IV, val av kalibratorkoncentrationer, kriterier för kalibrering samt statistisk inferens för variationskoefficienten.

6.1   Val av kalibratorkoncentrationer

Kalibratorernas koncentrationer bör väljas optimalt, enligt något kriterium. Många kriterier har föreslagits. Ofta minimeras något mått på variationen i skattningarna av kurvans parametrar. I artikel I föreslås att kriteriet istället baseras på felen i de inverst predikterade koncentrationerna. Generellt kan dessa fel uttryckas som en funktion av väntevärdet och variansen. Funktionen ska anta så litet värde som möjligt i hela mätområdet, vilket betyder att integralen av funktionen, från lägsta till högsta möjliga koncentration, ska minimeras.

Dessvärre beror väntevärdet och variansen i den inversa prediktionen på det sanna sambandet mellan respons och koncentration. Det sanna sambandet är okänt, och det varierar från tillfälle till tillfälle. Denna svårighet kan hanteras med Bayesianska metoder. Med en multivariat sannolikhetsfördelning beskrivs hur kurvans parametrar varierar slumpmässigt mellan tillfällen. Det är emellertid i praktiken svårt att veta vilken sannolikhetsfördelning parametrarna följer, och dessutom kräver metoden komplicerade beräkningar.

Alternativt kan kalibratorernas koncentrationer väljas så att de är optimala för ett enda fixerat samband mellan respons och koncentration, det vill säga för en enda given kalibreringskurva. Valet sägs i detta fall vara lokalt optimalt. Med denna metod bortses helt från det faktum att sambandet mellan respons och koncentration varierar mellan körningar.

I artikel I föreslås ett nytt tillvägagångssätt. Metoden för att bestämma lokalt optimala kalibratorkoncentrationer generaliseras. Varianser och väntevärden för inversa prediktioner behöver enligt den nya metoden inte nödvändigt beräknas betingat på en given kalibreringskurva, utan kan approximativt beräknas obetingat. Detta möjliggörs genom antaganden om kurvparametrarnas förväntade värden och varianser mellan tillfällen och med hjälp av linjär approximation.

6.2 Kriterier för kalibrering

Kalibreringskurvan anpassas vanligen till kalibratorernas responser med minsta kvadratmetoden. Eftersom variansen inte är konstant, utan ökar med responsens nivå, görs en vägd regression. Som vikter används inverterade skattningar av varianser. Vanligen skattas varianserna i förhand, i större datamaterial med mätvärden från många körningar.

I artikel II ges en möjlig förklaring till fenomenet att standardavvikelsen i immunkemi ofta är ungefärligt proportionell mot medelvärdet. Det kan vara den slumpmässiga variationen i de dispenserade volymerna av proven och kalibratorerna som orsakar en approximativt homogen variationskoefficient. Om detta är fallet, och om responsen är en funktion av antalet molekyler i den dispenserade volymen, är invers regression mer lämplig än vanlig regression.

Invers regression innebär att kalibreringskurvan anpassas, med minsta kvadratmetoden, till koncentrationer istället för till responser. Syftet är att minska medelkvadratfelen i de inversa prediktionerna. I artikel II föreslås att kalibreringskurvan anpassas med invers regression, och att inverterade kvadrerade kalibratorkoncentrationer används som vikter. Då minimeras kvadratsumman av alla relativa avvikelser i koncentration.

I artikel II ges också ett approximativt uttryck (4.1) för variansen i respons när variationskoefficienten är konstant i koncentration och sambandet mellan respons och koncentration beskrivs av en logistisk funktion med fyra parametrar (2.2).

6.3 Inferens för variationskoefficienten

Variationskoefficienten c är det vanligaste måttet på spridning i immunkemiska metoder. Den definieras som $c = s/m$, där s är standardavvikelsen och m medelvärdet (3.4), baserat på n observationer. Vanligen tas denna variationskoefficient c som skattning av populationsvariationskoefficienten γ, definierad enligt (3.5).

Variationskoefficienten används för att jämföra olika immunkemiska metoders precision och för att uttrycka krav på spridning i immunkemi. Ibland antas att variationskoefficienten är konstant i hela mätområdet. Ofta undersöks om variationskoefficienten är större under någon specifik omständighet än under någon annan. Av dessa anledningar behövs, vid analys av immunkemiska tekniker, statistisk inferens för variationskoefficienten. Det är särskilt vanligt att flera prov antas ha samma variationskoefficient och att denna så kallade gemensamma variationskoefficient skattas.

I denna avhandling utvecklas statistiska metoder för gemensamma variationskoefficienter i normalfördelade data. En skattningsfunktion tas fram, samt dess asymptotiska fördelning. Skattningsfunktionen är helt enkelt kvadratroten ur ett vägt medelvärde av de enskilda provens kvadrerade variationskoefficienter. Som vikter används antalet observationer minus 1. För små variationskoefficienter (<0.20) rekommenderas dessutom en justering för systematiskt fel, genom multiplikation med en faktor som beror på totala antalet observationer.

McKays approximation (3.10) för variationskoefficienten γ är approximativt χ^2-fördelad med $n-1$ frihetsgrader. Den kan användas för att beräkna approximativa konfidensintervall för variationskoefficienten. I artikel III föreslås ett nytt statistiskt test, baserat på McKays approximation, för hypotesen att två variationskoefficienter γ_1 och γ_2 är lika. Teststatistikan (4.2) ska jämföras med en F-fördelning. Enligt en simuleringsstudie fungerar det nya testet väl, särskilt med avseende på sannolikheten för fel av första slaget, även vid små stickprov. I artikel III föreslås dessutom metoder för att beräkna konfidensintervall för gemensamma variationskoefficienter, och pröva hypotesen att två eller flera gemensamma variationskoefficienter är lika. Även dessa metoder grundar sig på McKays approximation.

I artikel IV visas att McKays approximation är ickecentralt betafördelad, och att den dessutom är asymptotiskt normalfördelad med medelvärde $n-1$ och varians något mindre än $2(n-1)$. Eftersom variansen i en χ^2-fördelad variabel med $n-1$ frihetsgrader är $2(n-1)$ är, i denna mening, McKays approximation inte asymptotiskt korrekt. Avvikelsen bör emellertid i de flesta fall vara av försumbar betydelse. Baserat på resultaten i artikel IV ges i artikel III explicita uttryck för de två första momenten för McKays approximation.

References

Barndorff-Nielsen, O. E. (1986) Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika 73*, 307-322.

Barndorff-Nielsen, O. E. (1991) Modified signed log likelihood ratio. *Biometrika 78*, 557-563.

Belanger, B. A., Davidian, M. & Giltinan, D. M. (1996) The effect of variance function estimation on nonlinear calibration inference in immunoassay data. *Biometrics* 52, 158-175.

Bellio, R. (2003) Likelihood methods for controlled calibration. *Scandinavian Journal of Statistics* 30, 341-353.

Bennett, B. M. (1976) On an approximate test for homogeneity of coefficients of variation. In: Ziegler, W. J. (Ed.) Contributions to applied statistics dedicated to A. Linder, *Experientia Supplementum* 22, 169-171.

Bennett, B. M. (1977) LR tests for homogeneity of coefficients of variation in repeated samples. *Sankhyā,* Series B, 39, 400-405.

Benton, D., Krishnamoorthy, K. & Mathew, T. (2003) Inferences in multivariate-univariate calibration problems. *The Statistician* 52, 15-39.

Berkson, J. (1950) Are there two regressions? *Journal of the American Statistical Association* 45, 164-180.

Berkson, J. (1969) Estimation of a linear function for a calibration line: consideration of a recent proposal. *Technometrics* 11, 649-660.

Bhattacharya B. (2002) Tests of parameters of several gamma distributions with inequality restrictions. *Annals of the Institute of Mathematical Statistics* 54, 565-576.

Bhoj, D. S. & Ahsanullah, M. (1993) Testing equality of coefficients of variation of two populations. *Biometrical Journal* 35, 355-359.

Bonate, P. L. (2006) *Pharmacokinetic-pharmacodynamic modeling and simulation*. New York: Springer.

Bonett, D. G. & Seier, E. (2006) Confidence interval for a coefficient of dispersion in nonnormal distributions. *Biometrical Journal* 48, 144-148.

Bowman, K. O. & Shenton, L. R. (1983) Maximum likelihood estimators for the gamma distribution revisited. *Communications in Statistics – Simulation and Computation* 12, 697-710.

Box, G. E. & Cox, D. R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society,* Series B, 26, 211-252.

Brown, P. J. & Sundberg, R. (1987) Confidence and conflict in multivariate calibration. *Journal of the Royal Statistical Society*, Series B, 49, 46-57.

Brunnée, T., Seeberger, A, Kleine-Tebbe, J. & Kunkel, G. (1996) Comparison between two automated systems to determine specific IgE: CAP and ELItest. *Clinical and Experimental Allergy* 26, 1420-1427.

Carroll, R. J. (2003) Variances are not always nuisance parameters. *Biometrics* 59, 211-220.

Carroll, R. J. & Cline, D. B. H. (1988) An asymptotic theory for weighted least-squares with weights estimated by replication. *Biometrika* 75, 35-43.

Carroll, R. J. & Ruppert, D. (1984) Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association* 79, 321-328.

Carroll, R. J., Ruppert, D., Stefanski, L. A. (1995) *Measurement error in nonlinear models*. London: Chapman & Hall.

Chaloner, K. & Verdinelli, I. (1995) Bayesian experimental design: a review. *Statistical Science* 10, 237-304.

Cheng, C.-L. & Van Ness, J. W. (1999) *Statistical regression with measurement error*. London: Arnold.

Chow, S. C. & Liu J. P. (1995) *Statistical design and analysis in pharmaceutical science. Validation, process controls, and stability*. New York: Marcel Dekker.

Clinical and Laboratory Standards Institute (2004) Limits of detection and quantitation. Document EP17-A. Wayne, Pennsylvania: Clinical and Laboratory Standards Institute.

Cohen, A. C. & Whitten, B. J. (1982) Modified moment and maximum likelihood estimators for parameters of the three parameter gamma distribution. *Communications in Statistics − Simulation and Computation* 11, 197-216.

Cox, D. R. & Reid, N. (1987) Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society*, Series B, 49, 1-39.

David, F. N. (1949) Note on the application of Fisher's k-statistic. *Biometrika* 36, 383-393.

Davidian, M. (1990) Estimation of variance functions in assays with possibly unequal replication and nonnormal data. *Biometrics* 77, 43-54.

Davidian, M. & Carroll, R. J. (1987) Variance function estimation. *Journal of the American Statistical Association* 82, 1079-1091.

Davidian, M. & Giltinan, D. M. (1993) Some simple methods for estimating intraindividual variability in nonlinear mixed effects models, *Biometrics* 49, 59-73.

DeSilva, B., Smith, W., Weiner, R., Kelley, M., Smolec, J., Lee, B., Khan, M., Tacey, R., Hill, H. & Celniker, A. (2003) Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharmaceutical Research* 20, 1885-1900.

Dette, H. & Biedermann, S. (2003) Robust and efficient designs for the Michaelis-Menten model. *Journal of the American Statistical Association* 98, 679-686.

Dette, H., Melas, V. B. & Pepelyshev, A. (2003) Standardized maximin E-optimal designs for the Michaelis-Menten model. *Statistica Sinica* 13, 1147-1163.

Dette, H., Melas, V. B., and Wong, K. W. (2005) Optimal design for goodness-of-fit of the Michaelis-Menten enzyme kinetic function. *Journal of the American Statistical Association* 100, 1370-1381.

Doornbos, R. & Dijkstra, J. B. (1983) A multi sample test for the equality of coefficients of variation in normal populations. *Communications in Statistics - Simulation and Computation* 12, 147-158.

Engelhardt, M. & Bain, L. J. (1978) Construction of optimal unbiased inference procedures for the parameters of the gamma distribution. *Technometrics* 20, 485-489.

Feltz, C. J. & Miller G. E (1996) An asymptotic test for the equality of coefficients of variation from k populations. *Statistics in Medicine* 15, 647-658.

Fieller, E. C. (1932) A numerical test of the adequacy of A. T. McKay's approximation. *Journal of the Royal Statistical Society* 95, 699-702.

Fieller, E. C. (1954) Some problems in interval estimation. *Journal of the Royal Statistical Society*, Series B, 16, 175-185.

Finney, D. J. & Phillips, P. (1974) The form and estimation of a variance function, with particular reference to radioimmunoassay. *Applied Statistics* 26, 312-320.

François, N., Govaerts, B. & Boulanger, B. (2004) Optimal designs for inverse prediction in univariate nonlinear calibration models. *Chemometrics and Intelligent Laboratory Systems* 74, 283-292.

Fung W. K. & Tsang T. S. (1998) A simulation study comparing tests for the equality of coefficients of variation. *Statistics in Medicine* 17, 2003-2014.

Gerig, T. M. & Sen, A. R. (1980) MLE in two normal samples with equal but unknown population coefficients of variation. *Journal of the American Statistical Association* 75, 704-708.

Giltinan, D. M. & Davidian, M. (1994) Assays for recombinant proteins: a problem in non-linear calibration. *Statistics in Medicine* 13, 1165-1179.

Goos, P., Tack, L. & Vandebroek, M. (2001) Optimal designs for variance function estimation using sample variances. *Journal of Statistical Planning and Inference* 92, 233-252.

Gottschalk, P. G. & Dunn, J. R. (2005a) Determining the error of dose estimates and minimum and maximum acceptable concentrations from assays with nonlinear dose-response curves. *Computer Methods and Programs in Biomedicine*, 80, 204-215.

Gottschalk, P. G & Dunn, J. R. (2005b) The five-parameter logistic: a characterization and comparison with the four-parameter logistic. *Analytical Biochemistry* 343, 54-65.

Graf, U., Henning, H. J., Stange, K. & Wilrich, P. T. (1987) *Formeln und tabellen der angewandten mathematischen statistik*. Berlin: Springer.

Gupta, R. C. & Ma, S. (1996) Testing the equality of coefficients of variation in k normal populations. *Communications in Statistics - Theory and Methods* 25, 115-132.

Gut, A. (2005) *Probability: a graduate course*. New York: Springer.

Hage, D. S. (1999) Immunoassays. *Analytical Chemistry* 71, 294R-304R.

Hald, A. (1952) *Statistical theory with engineering applications*. New York: Wiley.

Hardin, R. H. & Sloane, N. J. A. (1993) A new approach to the construction of optimal designs. *Journal of Statistical Planning and Inference* 37, 339-369.

Hartley, H. O. (1950) The maximum *F*-ratio as a short cut test for heterogeneity of variance. *Biometrika* 37, 308-312.

Iglewicz, B. & Myers, R. H. (1970) Comparison of approximations to the percentage points of the sample coefficient of variation. *Technometrics* 12, 166-169.

Jones, G. & Rocke, D. M. (1999) Bootstrapping in controlled calibration experiments. *Technometrics* 41, 224-233.

Keating J., Glaser, R. E. & Ketchum, N. S. (1990) Testing hypotheses about the shape parameter of a gamma distribution. *Technometrics* 32, 67-82.

Krutchkoff, R. G. (1967) Classical and inverse regression methods of calibration. *Technometrics* 9, 425-439.

Krutchkoff, R. G. (1969) Classical and inverse regression methods of calibration in extrapolation. *Technometrics* 11, 605-608.

Lohrding, R. K. (1975) A two sample test of equality of coefficients of variation or relative errors. *Journal of Statistical Computation and Simulation* 4, 31-36.

McCullagh, P. & Nelder, J. A. (1989) *Generalized linear models.* 2nd ed. London: Chapman and Hall.

McKay, A. T. (1932) Distribution of the coefficient of variation and the extended 't' distribution. *Journal of the Royal Statistical Society* 95, 695-698.

Miller, E. G. & Karson, M. J. (1977) Testing equality of two coefficients of variation. *American Statistical Association: Proceedings of the Business and Economics Section*, Part I, 278-283.

Miller, G. E. (1991a) Asymptotic test statistics for coefficients of variation. *Communications in Statistics − Theory and Methods* 20, 3351-3363.

Miller, G. E. (1991b) Use of the squared ranks test to test for the equality of the coefficients of variation. *Communications in Statistics − Simulation and Computation* 20, 743-750.

Miller, G. E. & Feltz, C. J. (1997) Asymptotic inference for coefficients of variation. *Communications in Statistics - Theory and Methods* 26, 715-726.

Müller, P. (1999) Simulation-based optimal design. *Bayesian Statistics* 6, 459-474.

Nairy, K. S. & Rao, K. A. (2003) Tests of coefficients of variation of normal population. *Communications in Statistics - Simulation and Computation* 32, 641-661.

Nelson, L. S. (1987) Upper 10%, 5% and 1% points of the maximum *F*-ratio. *Journal of Quality Technology* 19, 165-167.

O'Connell, M., Belanger, B. A. & Haaland, P. D. (1993) Calibration and assay development using the four-parameter logistic model. *Chemometrics and Intelligent Laboratory Systems* 20, 97-114.

Osborne, C. (1991) Statistical calibration: a review. *International Statistical Review* 59, 309-336.

Owen, D. B. (1968) A survey of properties and applications of the noncentral t-distribution. *Technometrics* 10, 445-478.

Pang, W. K., Leung, P. K., Huang, W. K. & Liu, W. (2005) On interval estimation of the coefficient of variation for the three-parameter Weibull, lognormal and gamma distribution: A simulation-based approach. *European Journal of Operational Research* 164, 367-377.

Pawitan, Y. (2001) *In all likelihood: statistical modelling and inference using likelihood.* New York: Oxford University Press.

Pearson, E. S. (1932) Comparison of A. T. McKay's approximation with experimental sampling results. *Journal of the Royal Statistical Society* 95, 703-704.

Pitman, E. J. G. (1939) Tests of hypotheses concerning location and scale parameters. *Biometrika* 31, 200-215.

Pronzato, L. (2002) Information matrices with random regressors. Application to experimental design. *Journal of Statistical Planning and Inference* 108, 189-200.

Raab, G. M. (1981) Estimation of a variance function, with application to immunoassay. *Applied Statistics* 30, 32-40.

Rao, K. A. & Bhatt, A. R. S. (1995) Tests for Coefficient of Variation. *Journal of the Indian Society of Agricultural Statistics* 47, 225-229.

Rao, K. A. & Bhatta, A. R. S. (1989) A note on test for coefficient of variation. *Calcutta Statistical Association Bulletin* 38, 225-229.

Rao, K. A. & Vidja, R. (1992) On the performance of a test for coefficient of variation. *Calcutta Statistical Association Bulletin* 42, 87-95.

R Development Core Team (2008) R: a language and environment for statistical computing. Reference index. Version 2.7.2.

Reh, W. & Scheffler, B. (1996) Significance tests and confidence intervals for coefficients of variation. *Computational Statistics and Data Analysis* 22, 449-452.

Ricketts, J. H. & Head, G. A. (1999) A five-parameter logistic equation for investigating symmetry of curvature in baroreflex studies. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 277, R441-R454.

Rocke, D. M. & Jones, G. (1997) Optimal design for ELISA and other forms of immunoassay. *Technometrics* 39, 162-170.

Rocke, D. M. & Lorenzato, S. (1995) A two-component model for measurement error in analytical chemistry. *Technometrics* 37, 176-184.

Rodbard, D., Lenox, R. H., Wray, H. L. & Ramseth, D. (1976) Statistical analysis of radioligand assay data. *Methods of Enzymology* 37, 3-22.

Rodbard, D., Munson, P. J. & De Lean, A. (1974) Improved curve-fitting, parallelism testing, characterization of sensitivity and specificity, validation, and optimization for radioligand assays. In: Natelson, S., Pesce, A. J. & Dietz, A. (Eds.) *Radioimmunoassay and related procedures in medicine.* 469-504. Vienna: Int. Atomic Energy Agency.

Sadler, W. A. (2002) A new WIN32 computer program for estimating immunoassay variance functions. *Computer Methods and Programs in Biomedicine* 67, 195-199.

Sadler, W. A. & Smith, M. H. (1986) A reliable method of estimating the variance function in immunoassay. *Computational Statistics and Data Analysis* 3, 227–239.

Schwenke, J. R. & Milliken, G. A. (1991) On the calibration problem extended to nonlinear models. *Biometrics* 47, 563-574.

Seber, G. A. F. & Wild, C. J. (1989) *Nonlinear regression.* New York: Wiley.

Shafer, N. J. & Sullivan, J. A. (1986) A simulation study of a test for the equality of the coefficients of variation. *Communications in Statistics − Simulation and Computation* 15, 681-695.

Shalabh & Toutenburg, H. (2006) Consequences of departure from normality on the properties of calibration estimators. *Journal of Statistical Planning and Inference* 136, 4385-4396.

Sharma, K. K. & Krishna, H. (1994) Asymptotic sampling distribution of inverse coefficient-of-variation and its applications. *IEEE Transactions on Reliability* 43, 630-633.

Singh, M. (1993) Behaviour of sample coefficients of variation drawn from several distributions. *Sankhyā: The Indian Journal of Statistics*, Series B, 55, 65-76.

Tellinghuisen, J. (2000) Inverse vs. classical calibration for small data sets. *Fresenius Journal of Analytical Chemistry* 368, 585-588.

Tian L. (2005) Inferences on the common coefficient of variation. *Statistics in Medicine* 24, 2213-2220.

Tripathi, R. C, Gupta, R. C & Pair, R. K. (1993) Statistical tests involving several independent gamma distributions. *Annals of the Institute of Statistical Mathematics* 45, 773-786.

Umphrey, G. J. (1983) A comment on McKay's approximation for the coefficient of variation. *Communications in Statistics − Simulation and Computation* 12, 629-635.

Vangel, M. G. (1996) Confidence intervals for a normal coefficient of variation. *The American Statistician* 15, 21-26.

Verrill S. & Johnson R. A. (2007) Confidence bounds and hypothesis tests for normal distribution coefficients of variation. *Communications in Statistics - Theory and Methods* 36, 2187-2206.

Wagner, J. G. (1973) Properties of the Michaelis-Menten equation and its integrated form which are useful in pharmacokinetics. *Journal of Pharmacokinetics and Biopharmaceutics* 1, 103-121.

Warren, W. G. (1982) On the adequacy of the chi-squared approximation for the coefficient of variation. *Communications in Statistics − Simulation and Computation* 15, 21-26.

Wilson, C. A. & Payton, M. E. (2002) Modelling the coefficient of variation in factorial experiments. *Communications in Statistics − Theory and Methods* 31, 463-476.

Wong, A. C. M. (1992) Inferences on the shape parameter of a gamma distribution: a conditional approach. *Technometrics* 34, 348-351.

Wong, A. C. M. & Wu, J. (1998) Comparisons of approximate tail probabilities for the shape parameter of the gamma distribution. *Computational Statistics & Data Analysis* 27, 333-344.

Wong, A. C. M. & Wu, J. (2002) Small sample asymptotic inference for the coefficient of variation: normal and nonnormal models. *Journal of Statistical Planning and Inference* 104, 73-82.

Yanagimoto, T. (1988) The conditional maximum likelihood estimator of the shape parameter in the gamma distribution. *Metrika* 35, 161-175.

Yanagimoto, T. & Yamamoto, E. (1991) Constructing elementary procedures for inference of the gamma distribution. *Annals of the Institute of Statistical Mathematics* 43, 539-550.

Zeng, Q. & Davidian, M. (1997a) Testing homogeneity of intra-run variance parameters in immunoassay. *Statistics in Medicine* 16, 1765-1776.

Zeng, Q. & Davidian, M. (1997b) Bootstrap-adjusted calibration confidence intervals for immunoassay. *Journal of the American Statistical Association* 92, 278-290.

Zeng, Q. & Davidian, M. (1997c) Calibration inference based on multiple runs of an immunoassay. *Biometrics* 53, 1304-1317.

Zwanzig, S. (2000) On the criteria for experimental design in nonlinear error-in-variables models. In: Balakrishnan, N., Melas & V. B, Ermakov, S. (Eds.) *Advances in Stochastic Simulation Methods*. 153-164. Boston: Birkhäuser.

Acknowledgements

I thank my supervisor Dietrich von Rosen for continuous support. You have believed in me, encouraged me and always been available. Over and over again you have read my manuscripts and suggested improvements. You have let me 'learn by doing'.

I thank my friends and colleagues at the Department of Energy and Technology. I thank especially my fellow PhD-students in statistics: Zhanna, Idah and Saeid, and also Razaw, Jemila, Kristi, Karin, Geir and Anna. We have had a great time as PhD-students.

I thank all staff at the Unit of Applied Statistics and Mathematics for having kept in touch and for your interest in my work and research. And I thank all my new friends and co-workers at the Field Research Unit and the Department of Crop Production Ecology for providing a stimulating working environment and for motivating me to finish my studies.

The research was supported by the Centre of Biostochastics, Swedish University of Agricultural Sciences. Pharmacia Diagnostics AB sponsored the first years of my postgraduate education, and I thank Katarina Hedman for this support. I also thank my former colleagues at Pharmacia Diagnostics AB and Statisticon AB for pleasant company during part-time work.

My co-writers merit special thanks. I thank Steve Verrill for significant contributions and smooth cooperation, and Lars Söderström for valuable support and discussion.

I thank Edward Miller for kind interest and encouraging discussions.

Finally, I thank my beloved family. You give me, every day, the best recreation. Jenny, Viola and Vilhelm – I love you.