Mathematical Statistics
Stockholm University

Logistic regression in three-point designs

Mårten Vågerö and Rolf Sundberg

Research Report 2004:1

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

LOGISTIC REGRESSION IN THREE-POINT DESIGNS

Accepted for publication in
Encyclopedia of Biopharmaceutical Statistics, Second Edition.

Mårten Vågerö[1] and Rolf Sundberg[2]

[1] Department of Biostatistics, Clinical Science, AstraZeneca R&D Södertälje, S-151 85 Södertälje, Sweden. marten.vagero@astrazeneca.com
[2] Mathematical Statistics, Department of Mathematics, Stockholm University S-106 91 Stockholm, Sweden. rolfs@math.su.se

Key words: Tolerance threshold distribution, Maximum likelihood estimator, Separation, Infinite parameter estimates, D-optimal design, Dose-response, Normal approximation.

INTRODUCTION

Three-point designs are frequently used in phase II clinical studies to establish a dose-response relationship before proceeding to larger confirmatory studies. In cases with a binary outcome, a logistic regression model is often adopted. As these studies most often are of small or moderate size with little prior knowledge about the effective dose-interval, there is a substantial risk for an outcome with non-unique or infinite parameter estimates.

Morgan's book [1] discusses both sequential and fixed-point designs for estimation of a particular characteristic or the entire distribution of a dose tolerance threshold model for quantal response data. In a threshold model, each individual in the target population has a dose tolerance (threshold) to the exposure of a drug. If a subject (patient, healthy volunteer or animal) is administered a dose above its threshold, it will respond. Otherwise it will be a non-responder. If subjects are assumed to be randomly sampled from a logistic tolerance threshold distribution, the probability of responding when administered a dose x, can be written

$$P(x;\mu,\beta)=\frac{e^{\beta(x-\mu)}}{1+e^{\beta(x-\mu)}}, \qquad (1)$$

where μ is the median (ED50) and β represents the steepness of the dose-response curve. Parameter estimates that are asymptotically unbiased, and asymptotic standard errors for these estimates, can be determined by the method of maximum likelihood through many available statistical computer packages.

In small sample studies however, problems with infinite parameter estimates due to a complete or quasi-complete separation of the data may occur (see below or [2]). These situations also cause problems when working with some statistical computer packages. There are certainly textbooks discussing this problem, e.g. [3], but many modern textbooks on logistic regression do not mention it at all [4]. A published example where this was not properly addressed is discussed in [5]. In simple studies it is easily seen if there is a

complete or quasi-complete separation of the data leading to an infinite parameter estimate. However, in more complex models, with several explanatory variables, it need not be obvious from inspection of the outcome whether some parameter estimate is infinite or not. This paper will discuss the problem and practical implications of infinite parameter estimates in a three-point univariate logistic regression.

MAXIMUM LIKELIHOOD ANALYSIS

With the notation,

n_i = the number of subjects who are administered the dose x_i,

y_i = the number of subjects responding to dose x_i,

the log-likelihood derived from (1) is

$$L(\mu,\beta) = \sum_i y_i \beta(x_i - \mu) + \sum_i n_i \log\{1 - P(x_i;\mu,\beta)\} + \text{constant} \qquad (2)$$

The maximum likelihood estimator $(\hat{\mu}, \hat{\beta})$ is determined by maximising (2). The second derivatives form the observed information matrix, here denoted $J(\mu,\beta)$. We have

$$J(\hat{\mu},\ \hat{\beta}) = \begin{bmatrix} \hat{\beta}^2 \sum_i \cdot n_i \cdot \hat{\mathrm{P}}_i(1-\hat{\mathrm{P}}_i) & -\hat{\beta}\sum_i \cdot (x_i - \hat{\mu}) \cdot n_i \cdot \hat{\mathrm{P}}_i(1-\hat{\mathrm{P}}_i) \\ -\hat{\beta}\sum_i \cdot (x_i - \hat{\mu}) \cdot n_i \cdot \hat{\mathrm{P}}_i(1-\hat{\mathrm{P}}_i) & \sum_i (x_i - \hat{\mu})^2 \cdot n_i \cdot \hat{\mathrm{P}}_i(1-\hat{\mathrm{P}}_i) \end{bmatrix}, \qquad (3)$$

where $\hat{P}_i = P(x_i;\hat{\mu},\hat{\beta})$.

The variance-covariance matrix of the maximum likelihood estimator (MLE) is estimated by the inverse $J^{-1}(\hat{\mu},\hat{\beta})$ of the observed information matrix in the maximum likelihood point. We will use superscripts to denote the elements of J^{-1}, e.g. j^{11} for its upper left corner. The expected Fisher information, i.e. the expected value of J over the possible outcomes of data, will be denoted $I(\mu, \beta)$.

CHOICE OF DESIGN POINTS

Usually fixed-point designs divide the total number of subjects equally between all dose levels. The dose levels are ideally chosen symmetrically around the centre μ of the tolerance distribution and with an equal distance between the dose levels. See for example [6] or [1] for more details on design considerations for quantal response data. Consider now a three-point design, with 3n subjects equally divided between the three doses, $x_1 < x_2 < x_3$. The middle dose, x_2, is assumed located in the centre μ of the tolerance threshold distribution, and the other two doses, x_1 and x_3, symmetrically around x_2, say at $x_2 \pm \delta$. This makes P_2=1/2, and P_3=1-P_1, and the expected Fisher information, $I(\mu, \beta)$ is diagonal, thus $\hat{\mu}$ and $\hat{\beta}$ are approximately uncorrelated.

$$I(\mu,\beta) = n\begin{bmatrix} \beta^2(2P_1(1-P_1)+1/4) & 0 \\ 0 & 2P_1(1-P_1)\delta^2 \end{bmatrix} \qquad (4)$$

Many different optimality criteria are found in the literature. Most criteria are based on the information matrix and aim at minimising elements or functions of the elements in the inverse information. To achieve a μ-optimal design, I^{11} should be minimised, or equivalent $P_1(1\text{-}P_1)$ maximised, thus the design points should be chosen close to μ. An optimal design for estimating β should instead minimise I^{22}, that is maximise $P_1(1\text{-}P_1)\delta^2$. Using this optimality criteria, the spacing δ should be chosen so that δ=2.4β or equivalent so that P_1=1-P_3=0.083. A widely used optimality criterion is D-optimality, where the determinant of I is maximised [7]. This is a design with a good balance between a low risk of a separation in data and providing an informative dose-response relation. Using the D-optimality criteria, the design points are chosen so that P_1=1-P_3=0.136 or equivalently so that δ=1.85β.

DATA CONFIGURATIONS

In a binary dose-response model (with a single dose or a dose vector), some degenerate types of data configurations yield infinite and/or non-unique parameter estimates. Essentially following Albert and Anderson [2], we characterise the sample outcomes as follows:

1) *Only one response type represented in data.*
2) *A complete separation of responders from non-responders.*
2) *A quasi-complete separation of responders from non-responders.*
3) *An overlap of the two response types.*

These authors [2] discuss the problem of existence, finiteness and uniqueness of maximum likelihood estimates in the case of several explanatory variables. A further discussion of how to interpret and handle infinite parameter estimates in a general logistic regression model is found in [8]. Next we describe what Albert and Anderson's results imply in our case with only one explanatory variable, i.e. the dose.

Only One Response Type Represented in Data

The first situation is when the outcome consists of only responders or only non-responders. The only information provided by data in this trivial case is that the design points were chosen outside the effective dose range for the specific drug under investigation. The experiment must be supplemented with more design points in the direction of the effective dose range.

A Complete Separation of Responders from Non-responders

A complete separation means that there is a specific (but non-unique) dose level so that all responders are found at higher doses than the non-responders (or vice versa).
Thus,

$$\min_i (x_i | y_i > 0) > \max_i (x_i | y_i < n_i).$$

The probability is particularly high for obtaining such data when the dose-response curve rises steeply from ≈0 to ≈1 somewhere between $\max(x_i | y_i < n_i)$ and $\min(x_i | y_i > 0)$. It is easily verified that the maximum likelihood estimates are degenerated, being

$\hat{\beta} = +\infty$, and

$$\max_i (x_i | y_i < n_i) < \hat{\mu} < \min_i (x_i | y_i > 0)$$

A Quasi-complete Separation of Responders from Non-responders

A quasi-complete separation of responders from non-responders means that there is a unique dose level x^* in the design such that all responders occur at $x_i \geq x^*$, and all non-responders occur at dose levels $x_i \leq x^*$ (or vice versa), with both responders and non-responders at x^*.
In this situation, the highest probability for the observed data is obtained if the dose-response curve rises steeply at the particular dose x^*. Hence, the MLE in this case is

$(\hat{\beta}, \hat{\mu}) = (+\infty, x^*)$.

An Overlap of the Two Response Types

There is an overlap of the two response types if the smallest dose level with responders is smaller than the largest dose level with non-responders.
Under the parameterisation with μ and β, a degenerate MLE will occur in the extreme case when data suggest that there is no dose-response relation, i.e. $\hat{\beta} = 0$. In a three-point design, this means that the proportion of responders is the same for dose x_1 as for dose x_3. In this case,

$\hat{\beta} = 0$, and

$\hat{\mu}$ is non-unique or infinite.

Excluding this rare outcome, an overlap implies that the MLE $(\hat{\mu}, \hat{\beta})$ is unique and finite, and can be determined from the likelihood equations.

Properties of the Maximum Likelihood Estimator Illustrated

From a practical point of view, after having collected data from all patients in the experiment, we would like to base our inference on standard methods no matter what the underlying risk was for a separation of responders from non-responders.

If data suggest that there is a separation of responders from non-responders, there is not much information to gain from the experiment so you would need to supplement your experiment in some way. However, in the case all parameter estimates exist and are finite, you would like to determine the parameter estimates and make conclusions about the dose-response relation without taking into account the underlying risk of a separation of responders from non-responders.

Here we demonstrate that the asymptotic inference is valid as soon as $\hat{\beta}$ is finite. We illustrate this in a three-point design with a somewhat larger spacing than in a D-optimal design, design points at quantiles 10%, 50% and 90%, i.e. we chose x_i such that

$P_1 = 1\text{-}P_3 = 0.10$

$P_2 = ½$

This naturally represents an idealised situation, since in reality we do not know the position of the centre, nor the steepness. Placing the middle dose away from the centre would increase the risk of a complete or quasi-complete separation further stressing the importance of a good choice of design points.

The exact distribution for the MLE $(\hat{\beta}, \hat{\mu})$, together with its observed information matrix $J\left(\hat{\mu}, \hat{\beta}\right)$, was found by going through the $(n+1)^3$ possible distinct outcomes in a three-point design for a binary response variable with n observations per dose level.

Results

In a three-point design, $\hat{\beta} = +\infty$ if there are no responders at the lowest dose and no non-responders at the highest dose. The probability for this event is

$$P\left(\hat{\beta} = +\infty\right) \approx (1 - P_1)^n P_3^n . \qquad (5)$$

The corresponding probability for $\hat{\beta} = -\infty$ is negligible.

In the present investigation this formula yields the probabilities 0.122, 0.015 and 0.00003 for n=10, n=20 and n=50 observations per dose group respectively.

In Figure 1 the exact distribution of $\hat{\beta} - \beta$ standardised by j^{22} is presented for n=10 and n=20.

Figure 1: Empirical distribution of the $\sqrt{j^{22}}$-standardised error in $\hat{\beta}$ with n=10 (dot-dashed line) and 20 (short broken) observations per dose level. A $N(0,1)$ is included for reference (solid line)

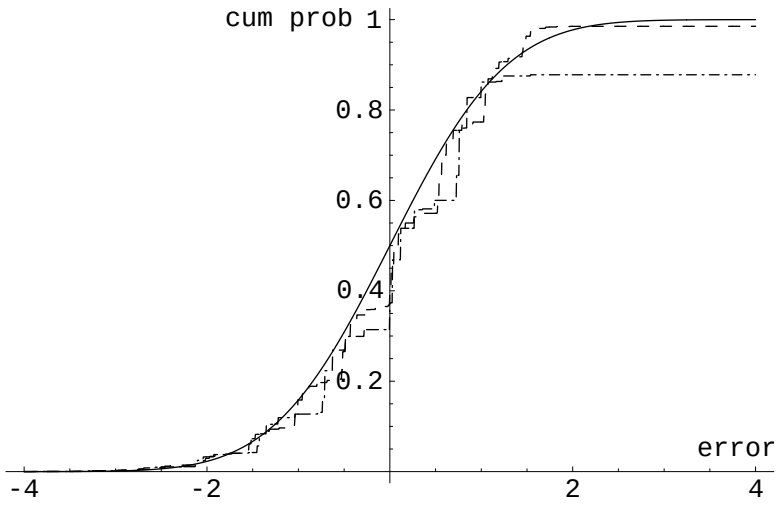

Table 1. Coverage probabilities for $-\infty, \hat{\beta} + z_\alpha \sqrt{j^{22}}$ intervals.

n	97.5% CI	95% CI	90% CI
10	0.9674	0.9592	0.8953
20	0.9663	0.9596	0.9061
50	0.9698	0.9530	0.8953

The j^{22}-standardised distribution curve of $\hat{\beta} - \beta$ follow a standard normal distribution quite well, indicating that the asymptotic inference about $\hat{\beta}$ is adequate, as soon as $\hat{\beta}$ is finite. In particular, the median of $\hat{\beta}$ is close to the true β-value. Hence, if centrality is measured by the median, there is no substantial systematic error in the MLE. Furthermore, as seen in Table 1, the coverage probabilities for one-sided confidence intervals $(-\infty, \hat{\beta} + z_\alpha \sqrt{j^{22}})$

based on a normal approximation are close to the nominal confidence levels for n=10, 20 and 50 even when $P(\hat{\beta}=+\infty)$ is substantial. Similar results where found for a D-optimal design and for designs deviating in location and spacing from the D-optimal, except that the probability for a degenerated MLE differed considerably. It should however be noted that the distribution of $(\hat{\beta}-\beta)/j^{22}$ has no or little probability mass in some intervals but takes large jumps in other points/intervals (see Figure 1). As the location of these intervals is different for different n-values, this causes an irregular pattern in the coverage probabilities seen in Table 2.

Table 2. Coverage probabilities for $\hat{\mu}\pm z_{\alpha/2}\sqrt{j^{11}}$ intervals, conditional on $\hat{\beta}<\infty$.

n	95% CI	90% CI	80% CI
10	0.9364	0.8782	0.7847
20	0.9474	0.8921	0.7920
50	0.9496	0.8982	0.7972

Figure 2. Empirical distribution of the $\sqrt{j^{11}}$-standardised error in $\hat{\mu}$ for the outcomes where $\hat{\beta} < +\infty$ with n=10 (dotted line). A $N(0,1)$ is included for reference (solid line).

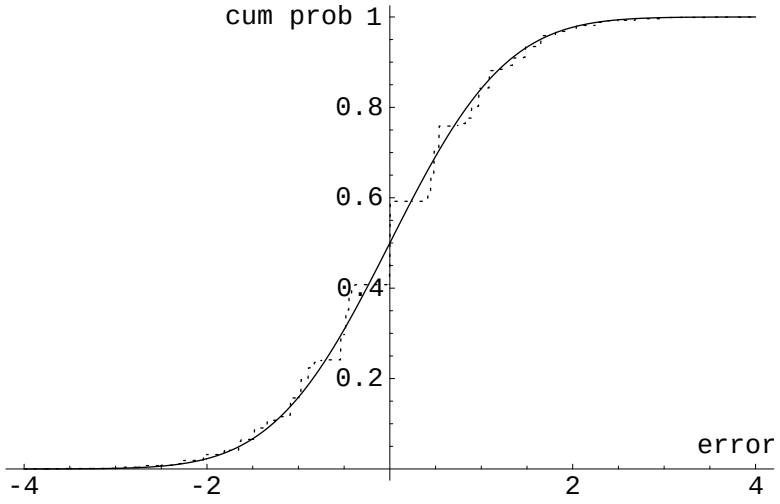

Figure 2 shows the $\sqrt{j^{11}}$-standardised error in $\hat{\mu}$ for n=10, given that $\hat{\beta}$ is finite. The figure illustrates that the $\sqrt{j^{11}}$-standardised error in $\hat{\mu}$ follows a standard normal distribution fairly well even with only n=10 observations per design point. For the case where n=20 observations per design point, the distribution of the error is even closer to the standard normal distribution. This case has been omitted in Figure 2 to increase readability. When $\hat{\beta} = +\infty$, use of the observed information from Eq. 3 would indicate a confidence interval for μ of length zero. However, in this case the likelihood is non-regular, having its maximum on the boundary of the parameter space, so the observed information in Eq. 3 is not justified and should not be used. The results from the exact calculations presented in Table 2 demonstrate that the conditional coverage probabilities are fairly close to the nominal level. It must be stressed that these coverage probabilities are conditional on a finite $\hat{\beta}$, indicating that as soon as we have an overlap of the two response types, asymptotic normality of $\hat{\mu}$ seems to be appropriate. Qualitatively the same results have been established not only for a D-optimal design, but also for

designs deviating in location and spacing from the D-optimal, e.g. a design in [5] with an extremely large spacing yielding a finite $\hat{\beta}$ with only probability 0.3. A partial explanation of this remarkable fact is that $\hat{\mu}$ and $\hat{\beta}$ are approximately uncorrelated (note that the expected fisher information in this case is diagonal, see Eq. 4).

DISCUSSION

The overall conclusion of this work is that three-point designs do not produce systematic errors in the distribution of the maximum likelihood estimator of the steepness parameter β, but if the design-points are too widely spaced the probability of an infinite $\hat{\beta}$ is unreasonably high. If one succeeds in choosing the spacing according to a D-optimal design, $\hat{\beta}$ seems to be an effective estimator with a reasonably low probability of being infinite, even with only 10 subjects per dose level. However, from a practical point out view it is difficult to determine the doses D-optimally, since it must be based on prior beliefs about the dose-response relationship.

Prior beliefs are required also for a Bayesian analysis of data. Using a data augmentation prior (DAP) (see Ch. 13 in [4]) would be equivalent with adding some extra observations to the experiment. This would reduce/eliminate the risk for infinite or non-unique parameter estimates, but it might then hide the lack of information in the data themselves.

An important practical conclusion of the examples investigated is that once we have an outcome yielding a finite $\hat{\beta}$, asymptotic normality-based confidence intervals are justified for inference about μ given the finite $\hat{\beta}$ and with some extra care also for inference about β.

Therefore, the main problem is outcomes with infinite $\hat{\beta}$. Most commonly used statistical computer packages provide warnings that we might have the case of a complete or quasi-complete separation, but some of them nevertheless return finite estimates for all parameters in these situations. It is also important when teaching logistic regression to users of these statistical tools, to discuss the interpretation of a complete or quasi-complete separation of the data.

In practice, what should be done with an experiment that gives an infinite $\hat{\beta}$? First of all, its data indicate that the steepness is greater than expected in advance. The next step would either be to stop the trial here, or more likely, to continue the experiment by including more subjects and more design points in the trial. Different sequential types of procedures have been suggested (see [1]).

CONCLUSIONS

It is concluded that the main problem in three-point designs with a binary outcome is the risk of observing an infinite $\hat{\beta}$. We need to take this risk into account when designing the experiment. A D-optimal design provides an informative dose-response relation and still has a low risk of observing an infinite $\hat{\beta}$. However, irrespectively of the design, once we have an outcome where all parameter estimates exist and are finite, asymptotic inference is remarkably adequate even in small-sample studies.

REFERENCES

1. Morgan B. J. T. (1992). *Analysis of Quantal Response Data*. Chapman & Hall, London.
2. Albert A., Anderson J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* 71: 1-10.
3. Hosmer D, Lemeshow S (1989). *Applied logistic regression*. Wiley, New York.
4. Christensen R (1997). *Log-linear models and logistic regression*, 2nd ed., Springer-Verlag, New York.
5. Vågerö M. (2000). A remark on small-sample properties of logistic regression in three-point designs. *J. Biopharm. Stat. 10: 573-587.*
6. Finney D.J. (1978). *Statistical Methods in Biological Assay*, 3rd ed. Griffin, London.
7. Kalish L. A. (1990). Efficient Design for Estimation of Median lethal Dose and Quantal Dose-Response Curves. *Biometrics* 46: 737-748.
8. Kolassa J. E. (1997). Infinite Parameter Estimates in Logistic Regression, with Application to Approximate Conditional Inference. *Scand. J. Statist.* 24: 520-530.