STATISTICS IN MEDICINE, VOL. 14, 811–819 (1995)

THE LOG TRANSFORMATION IS SPECIAL

OLIVER N. KEENE

Department of Medical Statistics, Glaxo Research and Development Ltd., Greenford Road, Greenford, Middlesex, UB6 0HE, U.K.

SUMMARY

The logarithmic (log) transformation is a simple yet controversial step in the analysis of positive continuous data measured on an interval scale. Situations where a log transformation is indicated will be reviewed. This paper contends that the log transformation should not be classed with other transformations as it has particular advantages. Problems with using the data themselves to decide whether or not to transform will be discussed. It is recommended that log transformed analyses should frequently be preferred to untransformed analyses and that careful consideration should be given to use of a log transformation at the protocol design stage.

1. INTRODUCTION

The use of *t*-tests, analysis of variance and analysis of covariance for continuous positive data on an interval scale is widespread. One of the easiest modifications to these simple parametric methods is the prior use of a log transformation.

Conventional wisdom dictates that the data should be analysed untransformed and the residuals examined for outliers, deviations from Normality and other indications of departures from the required assumptions. If this investigation indicates that the problems are severe then transformations may be considered.[1] Often it is recommended that a Box–Cox analysis be performed.[2] These procedures may lead to a log transformation, but may equally well lead to some other transformation and depend on the actual data observed.

In clinical trials, the analysis strategy should as far as possible be specified in advance in the protocol. Because many of the approaches to decisions on transformations are essentially subjective, this has led to a widespread suspicion of the use of any transformation.

Trials performed by pharmaceutical companies are heavily influenced by the attitudes of regulatory authorities. This suspicion of transformations is reflected in the FDA guideline[3] for the format and content of the statistical section of an application. This states:

> 'Unnecessary data transformation should be avoided. In the event a data transformation was performed, a rationale for the choice of data transformation along with interpretation of the estimates of treatment effects based on transformed data should be provided.'

At the least, this provides some discouragement to a pharmaceutical company to transform their data. It is clear that an industry statistician should not analyse the data using a number of transformations and pick the most favourable to the company. However, a consequence of this guideline is that the log transformation is grouped with all other types of transformation and is given no special status.

CCC 0277–6715/95/080811–09

Received March 1993
Revised April 1994

Table I. Gastric half-emptying time (min)

Subject	Treatment A		Treatment B		Treatment C	
	Period	Response	Period	Response	Period	Response
1	1	84	2	62	3	58
2	3	87	2	108	1	38
3	3	85	1	85	2	96
4	1	82	3	46	2	61
5	2	83	1	70	3	46
6	2	110	3	110	1	66
7	3	215	2	86	1	42
8	1	50	3	46	2	34
9	2	92	3	50	1	80
10	1	70	2	61	3	55
11	3	97	1	40	2	78
12	2	95	1	147	3	57
Mean		96		76		59
SD		40		33		19

1.1. Example: gastric emptying study

This was a three period, three treatment crossover study in 12 volunteers. The primary endpoint was gastric half-emptying time (min) and the data are given in Table I. In many medical journals, data such as these would be analysed untransformed. For the comparison of treatments A and B a non-significant P-value ($P = 0{\cdot}14$) would be quoted and possibly a confidence interval for the treatment difference (95 per cent CI: -7 to 47 min). There might be some discussion of the large value for subject 7 on treatment 1 and even an additional analysis excluding this value.

The increase in standard deviation with the mean is suggestive of the need for a log transformation. If the data are log transformed prior to analysis, the increase of 29 per cent between treatments A and B now approaches significance ($P = 0{\cdot}083$; 95 per cent CI: -4 per cent to 72 per cent). The large value for subject 7 is no longer a potential outlier.

The log transformed analysis is more supportive of a treatment effect than the untransformed analysis. However, no log transformation is specified in the protocol. Given the suspicion of log transformed analyses, it is clearly easier to convince a sceptical reviewer of the possibility of a treatment effect if the log transformed analysis had been planned in advance.

2. WHY DO WE NEED TO TRANSFORM CONTINUOUS DATA?

There are a number of reasons why an analysis on a ratio scale may be preferred to an analysis on the original scale.

2.1. Clinical importance relates to a ratio scale

When the magnitude of an effect is commonly perceived in terms of percentage change between treatments, this is usually a good indication that the clinical importance relates to a ratio scale. It seems perverse to base the statistical analysis on absolute values when changes to small responses are more clinically important than changes to large responses. Where baseline information is available, a common approach is to analyse the percentage change of a variable from baseline. Patients with small baseline values can have a greatly inflated influence on the analysis of

percentage change and this is generally a poor way of incorporating baseline information.[4] A log transformation weights observations automatically according to a ratio scale and reduces problems associated with percentage changes from baseline.

2.2. End-point represents a ratio of variables or a reciprocal

In general, if two variables are approximately Normally distributed with similar variability, then it is unlikely that their ratio will have an approximate Normal distribution. Kronmal[5] discusses other problems associated with use of untransformed ratios as the dependent variable in regression analyses. By applying a log transformation, the ratio of the variables is now expressed as a difference of two variables and the assumptions required by analysis of variance or regression analysis are usually much more realistic.[6] It is sometimes arbitrary which way round a ratio is expressed and an analysis of the log of the ratio makes this irrelevant. Where a variable is the reciprocal of another, a log transformation allows identical inferences for both variables.

2.3. Inference depends on ratios

Often the inference to be made depends on ratios, for example, conclusions of bioequivalence depend on ratios of treatments. An analysis based on untransformed data then requires division of the treatment difference by an estimated treatment mean. Either this is done crudely by ignoring the estimation error in the treatment mean or more precisely by application of Fieller's theorem.[7] Derivation of a confidence interval with finite positive limits is then only possible provided the estimates of both treatment means are significantly greater than zero.[8]

A much more straightforward solution is provided by use of a log transformation, where the treatment differences will automatically provide treatment ratios when transformed back to the original scale.

2.4. Multiplicative models

Effects frequently act multiplicatively and variability often increases with the size of the measurement.[9] A log transformation is explicitly recommended when the standard deviation is proportional to the mean value.[10]

Variables such as biochemical measurements typically show a skewed distribution,[11] which can often be made symmetric using a log transformation. It has been argued that 'the theoretical justification for using this transformation for most scientific observations is probably better than that for using no transformation at all'.[12]

2.5. Example: pharmacokinetic studies

The need for a log transformation of AUC and Cmax in pharmacokinetic studies has been discussed for a long time. These discussions illustrate many of the above points.

The inference to be made is for ratios between treatments. Pharmacokinetic considerations indicate that effects act multiplicatively. In simple situations, AUC is inversely related to plasma clearance and Cmax is inversely related to the volume of distribution.[13] Recent consensus statements[14] and regulatory guidelines[15, 16] have unequivocably favoured the prior use of log transformation. All these documents recommend that the data are not used to determine the correct transformation.

Another parameter commonly determined is the half-life ($t_{1/2}$), which is calculated according to the following formula:

$$t_{1/2} = \ln(2)/k_e$$

where k_e is the elimination rate constant. Use of a log transformation for $t_{1/2}$ yields identical inferences for $t_{1/2}$ and k_e.

3. STATISTICAL METHODS

3.1. Statistical objectives

The objectives of a clinical trial are typically to compare two or more treatments and to provide estimates and confidence intervals for the size of the effect as well as *P*-values. These should be made on a meaningful clinical scale. This is now required by EC GCP[17] and confidence intervals are explicitly requested by some medical journals.[18] The usefulness of analyses should therefore be guided by their ability to produce such estimates readily.

Analyses should take full account of the experimental design and be able to identify outlying values and interactions where these are of interest.

3.2. Families of transformations

The most well-known family of transformations is the Box–Cox:[19]

$$z = \begin{cases} (y^\lambda - 1)/\lambda & (\lambda \neq 0) \\ \log(y) & (\lambda = 0) \end{cases}$$

where $\lambda = 1$ implies no transformation, $\lambda = 0$ gives a log transformation, $\lambda = 0{\cdot}5$ a square root and $\lambda = -1$ a reciprocal.

As a referee has pointed out, the log is the only member of the Box–Cox family of transformations for which the transform of a positive-valued variable can be truly Normal, because the transformed variable is defined over the whole of the range from $-\infty$ to ∞.

A modified version of the log transformation may be obtained by using the transformation $z = \log(y - c)$, where c corresponds to a lower bound for y.[20] Berry[21] has proposed that this transform be used 'whenever a parametric analysis is planned'. One of the motivations behind this is that a log transformation may give undue weight to small values.

While a particular transformation may satisfy statistical criteria regarding distributional assumptions, there is a compelling reason to favour the log transformation. Treatment means may be directly transformed back to the original scale for all these transformations, but for treatment differences only the simple log transformation provides a direct back transformation, allowing treatment differences on the transformed scale to be interpreted as ratios on the original scale. If a transformation is repeatedly and routinely used for the same type of data, then some familiarity with the interpretation of treatment differences on the transformed scale may arise with experience, but most of these transformations are used for a particular dataset and a different approach will often be followed for a subsequent dataset.

3.3. Generalized linear models

While analysis of variance assumes a model where explanatory variables produce additive effects on the response and where the error variance is constant, generalized linear models split the

model into systematic and random components. Transformations seek to achieve both objectives simultaneously. An advantage of generalized linear models over simple data transformation is that a transformation to produce additivity can be made quite independently of a transformation to produce approximate Normality or constancy of variance.[22] Extensions of this approach allow the error component to be modelled as a function of parameters.

Their main disadvantage is the analysis of clinical trial data lies in their more complex nature and consequent unfamiliarity to the non-statistical audience. Diagnostic tools are also less well-developed compared with a least squares analysis of log transformed data.[23]

A log transformation is still special in the framework of generalized linear models in that only the identify and log link functions allow the simple interpretation of treatment differences as discussed above.

3.4. Non-parametric methods

Non-parametric methods (or distribution free methods) are not as susceptible as model based methods to controversy over the justification of assumptions.[24] They are also often presented as an alternative for situations when an untransformed parametric analysis appears unsatisfactory.[2]

It is clear that the Wilcoxon approach is useful for analyses of two-treatment studies, whether crossover or parallel group. Methods for evaluating confidence intervals based on a Wilcoxon approach are now in widespread use.[25,26]

Some authors advocate use of the rank transformation, which may be useful for minimizing the impact of outliers and in deriving *P*-values. Methods of deriving estimates of treatment effects from such analyses have not been widely discussed. In general, estimation based on non-parametric methods works well for the simple cases, but where the design is more complex, for example involving covariates, more research is required.

In clinical studies, one of the advantages of non-parametric methods, that they are robust to outliers, may be a disadvantage because such points may reflect a sub-population which requires investigation.[27]

For some common designs, for example crossovers with more than two periods, no non-parametric method is recommended by a recent textbook.[28] Assessment of interactions, such as treatment by centre interactions, typically requires use of parametric methods.[24]

In the situations where a non-parametric analysis provides a good alternative to parametric analysis the issue of use of a transformation is still important. For two period crossovers the standard analysis[29] is based on period differences, which will not in general provide the same analysis as one based on an analysis of ratios between periods. The standard method of calculation for confidence intervals corresponding to the Wilcoxon two sample case is based on individual treatment differences.[25] Again a transformation to ratios will affect the confidence intervals.

4. LET THE DATA DECIDE?

4.1. Procedure

A common method of data analysis frequently recommended in books on statistics,[2,30] uses a procedure which will be called 'Let the data decide'. This approach requires the data first to be analysed untransformed and then an assessment of goodness-of-fit to be made. If there is evidence of a departure from assumptions, a choice from a variety of transformations or non-parametric methods is made.

4.2. Problems with let the data decide

Although the 'let the data decide' rule appears appealing for one-off data sets in uncontroversial areas, problems arise in its practical use to decide the scale of measurement to report treatment effects in the analysis of clinical trials:

(a) Problems with decision procedures – many studies are just not large enough to distinguish between an untransformed and log transformed analysis. Buck[31] simulated Normal and log-Normal data for a two period crossover. His work indicates that with a coefficient of variation of 20 per cent and 32 subjects, the correct transformation will be chosen by a Box–Cox likelihood criteria in only 70 per cent of cases.

(b) Multiple testing dilemma – the 'let the data decide' approach is a two-stage procedure in the analysis. Should the uncertainty over choice of transformation be allowed for in confidence intervals and *P*-values? The suspicion remains that the statistician has selected the most favourable analysis.

(c) Consistency – many analyses form part of a series of trials, for example a package of studies with a new drug submitted to a regulatory authority. Use of a 'let the data decide' rule will often lead to inconsistent decisions across studies, since these rules will sometimes choose the wrong transformation and sometimes there will be a slight change in the distribution. It is clearly desirable that estimates of effect are comparable from one trial to another.

A similar but more immediate problem can occur within a single trial. Often measurements are made and analysed at more than one time point or on more than one population. It is quite conceivable that these rules, particularly if well-defined, could decide for transformation at one time point and against transformation at another. Clearly no analyst would wish to adopt different transformations at different timepoints.

When the assumptions of a simple parametric analysis are not justified, then an analysis should be sought which retains the desired scale for reporting treatment differences. Modelling of the error structure, using generalized linear models (Section 3.3) or non-parametric analysis (Section 3.4) fulfils this requirement.

4.3. Example: cortisol suppression

This was a double-blind, parallel group study with five treatments and 50 subjects. Measurements of plasma cortisol were made over 3 days: for one day prior to treatment; for 24 hours after one dose, and 24 hours after a second dose. The endpoints to be compared between treatments were mean cortisol over the 24 hour periods on day 2 and day 3. The planned analysis was analysis of covariance using the day 1 mean as a covariate. The data are shown in Table II.

If the cortisol is analysed untransformed, plots of standardized residuals against predicted values are hard to interpret (see Figure 1 for day 3 values). A Box–Cox analysis is similarly inconclusive:

Day 2 Estimate for $\lambda = 0{\cdot}2$, 95 per cent CI $= (-1{\cdot}0, 1{\cdot}3)$

Day 3 Estimate for $\lambda = 0{\cdot}5$, 95 per cent CI $= (-0{\cdot}1, 1{\cdot}1)$.

Some would now argue that a square root transformation is indicated or that a $\log(y - c)$ transformation should be applied, but this leaves the problem of estimating the relative cortisol suppression of treatments and most importantly the confidence intervals.

For this study, a log transformation for cortisol was pre-specified in the protocol. This was decided because the clinical importance of changes in cortisol are conventionally viewed in multiplicative terms and the inference to be made concerned the per cent fall in cortisol between

Table II. Plasma cortisol (nmol/l)

Treatment	Subject	Day 1	Day 2	Day 3
A	1	220·6	192·6	186·6
	6	151·2	150·9	117·2
	11	164·6	156·9	145·7
	19	191·4	160·0	147·2
	22	224·6	197·9	181·3
	23	237·1	189·9	154·1
	32	237·8	230·3	236·0
	38	254·8	171·0	176·4
	47	161·9	118·0	127·9
	50	165·4	141·7	137·4
B	2	189·6	178·9	141·1
	4	196·6	175·4	168·2
	13	185·3	169·2	136·7
	17	190·9	168·9	144·2
	24	218·9	181·6	176·9
	25	170·6	159·8	125·1
	37	218·4	206·2	176·4
	40	244·0	206·0	229·7
	45	306·7	190·8	172·6
	46	178·8	149·1	103·2
C	8	198·0	143·0	121·6
	10	174·4	159·5	130·2
	14	194·2	168·7	71·6
	20	160·8	156·9	135·6
	29	213·3	148·4	127·4
	30	138·6	134·9	84·6
	34	151·5	153·0	129·1
	39	200·6	189·4	161·2
	41	271·8	219·9	178·2
	49	181·4	163·2	159·0
D	7	144·7	160·2	185·0
	9	193·7	159·3	148·9
	15	169·4	199·7	182·8
	18	177·0	127·6	128·6
	27	252·5	221·4	194·1
	28	220·8	224·2	191·7
	31	135·6	150·9	176·9
	36	141·2	146·7	192·7
	44	209·1	201·9	304·8
	48	179·1	164·2	162·0
E	3	219·0	211·7	200·6
	5	222·3	261·5	239·8
	12	199·9	200·1	220·9
	16	151·4	160·7	163·1
	21	135·2	149·2	157·5
	26	238·0	259·8	241·7
	33	188·1	225·5	197·6
	35	189·3	195·0	199·7
	42	206·7	203·1	223·1
	43	148·1	186·4	188·5

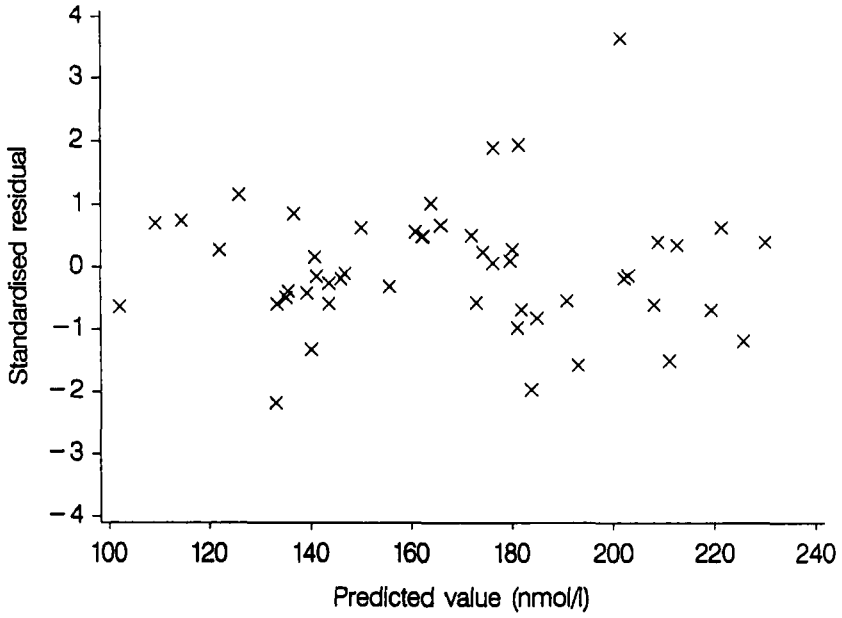

Figure 1. Plot of standardized residuals against predicted values for day 3 mean cortisol

treatments. The log transformed analysis provides a reasonable (if not perfect) fit to the data. Therefore one should feel comfortable presenting the log transformed analysis as the main analysis.

5. RECOMMENDATIONS

For continuous positive data measured on an interval scale, a log transformed analysis should frequently be preferred to an untransformed analysis. No special justification beyond that sufficient to support an untransformed analysis should be required from the data obtained. Unnecessary use of other transformations should be avoided.

If the use of a log transformation is chosen on a case-by-case basis then this will lead to inconsistencies and sometimes the wrong choice will be made. Prior considerations of the endpoints themselves and of the inference to be made will often indicate whether a ratio scale for reporting treatment effects will be appropriate. In these cases, protocols should specify the expected use or lack of use of a log transformation (or log link function). Additional modelling of the error structure or non-parametric analysis should be performed when the assumptions of the simple analysis are not justified or where previous experience has indicated that this is necessary.

ACKNOWLEDGEMENT

I would like to thank Alan Ebbutt for his encouragement and for suggesting the title of this paper.

REFERENCES

1. Berry, D. A. 'Basic principles in designing and analyzing clinical studies', *in* Berry, D. A. (ed.) *Statistical Methodology in the Pharmaceutical Sciences*, Marcel Dekker, New York, 1990.
2. Chinchilli, V. M. 'Clinical efficacy trials with quantitative data', *in* Peace, K. E. (ed.), *Biopharmaceutical Statistics for Drug Development*, Marcel Dekker, New York, 1988.

3. *Guideline for the Format and Content of the Clinical and Statistical Sections of New Drug Applications*, Food and Drug Administration, Public Health Service, US Department of Health and Human Services, 1988.
4. Senn, S. J. 'Letter to the editor', *Statistics in Medicine*, **10**, 1157–1160 (1991).
5. Kronmal, R. A. 'Spurious correlation and the fallacy of the ratio standard revisited', *Journal of the Royal Statistical Society, Series A*, **156**, 379–392 (1993).
6. Senn, S. J. 'The use of baselines in clinical trials in bronchodilators', *Statistics in Medicine*, **8**, 1339–1350 (1989).
7. Locke, C. S. 'An exact confidence interval from untransformed data for the ratio of two formulation means', *Journal of Pharmacokinetics and Biopharmaceutics*, **12**, 649–655 (1984).
8. Chow, S-C. and Liu, J-P. *Design and Analysis of Bioavailability and Bioequivalence Studies*, Marcel Dekker, New York, 1992.
9. Oldham, P. D. *Measurement in Medicine*, The English Universities Press Ltd., London, 1968.
10. Draper, N. and Smith, S. *Applied Regression Analysis*, 2nd edn., Wiley, New York, 1981.
11. Flynn, F. V., Piper, K. A. J., Garcia-Webb, P., McPherson, K. and Healy, M. J. R. 'The frequency distributions of commonly determined blood constituents in healthy blood donors', *Clinica Chimica Acta*, **52**, 163–171 (1974).
12. Gaddum, J. H. 'Lognormal distributions', *Nature*, **156**, 463 (1945).
13. Westlake, W. J. 'Bioavailability and bioequivalence of pharmaceutical formulations', *in* Peace, K. E. (ed.), *Biopharmaceutical Statistics for Drug Development*, Marcel Dekker, New York, 1988.
14. Cartwright, A. C. 'The March 1991 Barcelona Drug Information Association international consensus statement on bioavailability/bioequivalence testing', *Drug Information Journal*, **25**, 473–482 (1991).
15. *Guidance: Statistical Procedures for Bioequivalence Studies using a Standard Two-treatment Crossover Design*, Office of Generic Drugs, Food and Drug Administration, Public Health Service, US Department of Health and Human Services, July 1992.
16. *CPMP Guideline: Investigation of Bioavailability and Bioequivalence*, Commission of the European Communities 111/54/89-EN, 1992.
17. *Good Clinical Practice for Trials on Medicinal Products in the European Community*, Commission of the European Communities 111/3976/88-EN, 1990.
18. Gardner, M. J., Machin, D. and Campbell, M. J. 'Use of check lists in assessing the statistical content of medical studies', *British Medical Journal*, **292**, 810–812 (1986).
19. Box, G. E. P. and Cox, D. R. 'An analysis of transformations', *Journal of the Royal Statistical Society, Series B*, **26**, 211–252 (1964).
20. Aitchison, J. and Brown, J. A. C. *The Lognormal Distribution*, Cambridge University Press, 1957.
21. Berry, D. A. 'Logarithmic transformations in ANOVA', *Biometrics*, **43**, 439–456 (1987).
22. McCullagh, P. and Nelder, J. A. *Generalized Linear Models*, 2nd edn., Chapman and Hall, London, 1989.
23. Firth, D. 'Multiplicative errors: log-normal or gamma?', *Journal of the Royal Statistical Society, Series B*, **50**, 266–268 (1988).
24. Koch, G. G. and Sollecito, W. A. 'Statistical considerations in the design, analysis, and interpretation of comparative clinical trials', *Drug Information Journal*, **18**, 131–151 (1984).
25. Conover, W. J. *Practical Non-parametric Statistics*, 2nd edn., Wiley, New York, 1980.
26. Hauschke, D., Steinijans, W. J. and Diletti, E. 'A distribution-free procedure for the statistical analysis of bioequivalence studies', *International Journal of Clinical Pharmacology, Therapy and Toxicology*, **28**, 72–78 (1990).
27. Shuirmann, D. J. 'Discussion to session on outliers', *Proceedings of Bio-international '89*, Toronto, Canada, October 1–4, 184 (1989).
28. Jones, B. and Kenward, M. G. *Design and Analysis of Cross-over Trials*, Chapman and Hall, Great Britain, 1989.
29. Koch, G. G. 'The use of non-parametric methods in the statistical analysis of the two period changeover design', *Biometrics*, **28**, 577–584 (1972).
30. Fisher, L. D. and Van Belle, G. V. *Biostatistics*, Wiley, New York, 1993.
31. Buck, R. 'Discussion to session on log transformation', *Proceedings of Bio-international '89*, Toronto, Canada, October 1–4, 165–172 (1989).