Statistical considerations for the mouse LD50 assay and the impact of toxin reference standards

Rose Gaines Das

National Institute for Biological Standards and Control

Proposed use of LD50 (1)

- Toxicological
  - Classification of toxicity
    - Optimal adaptive designs for acute oral toxicity assessment, N. Stallard (2006), Journal of Statistical Planning and inference 136:1781-1799
  - Detection of presence / absence
    - At a defined 'limit' concentration
    - Using a defined method
  - In these cases precise estimation of the LD50 is not required

Proposed use of LD50 (2)
Use of LD50 to define potency

- "The dose of botulinum A toxin preparations is expressed in terms of units. The manufacturers state that one unit corresponds to the median lethal dose (LD50) injected intraperitoneally into mice under defined conditions."
- (Martindale 32nd Edn)

Proposed use of LD50 (2)

- Determination of the potency of a therapeutic / medicinal product
  - Therapeutic products must be safe and effective
  - Thus accurate and precise determination of potency is required
  - Regulatory authorities set limits on required precision of estimated potency

Estimation of LD50

- Estimation requires some assumptions
- Statistical assumptions: Model
  - Simplest model is that the proportion of animals responding increases with increasing concentration of toxin
  - Typically stronger assumptions about the dose – response relation, tolerance distribution
- Summary: Finney DJ (1985) The median lethal dose and its estimation (Arch. Toxicol. 56: 215-218)

Estimation of LD50 Requires Data

Experimental Design: Method of Data Collection

- Selection of number and 'concentration' of dose levels
- Total number of subjects and distribution among dose levels
- Optimal designs are parameter dependent
  - Thus, if the 'true' LD50 is known, an optimal design can be constructed.
- Depend on aims of experiment

Experimental Designs for Estimation of LD50

- Design must depend on local facilities / capabilities as well as available prior information
- Two broad principles for LD50 estimation
  - Doses with response levels nearer 50% are more informative than doses with more extreme responses
  - The greater the number of independent replicates, the greater the precision of the estimate

Statistical considerations

Precision related to design

Assume probit model, LD50=1, slope=5

Statistical Weight (Reciprocal of variance of log10 of LD50)		Number of mice	
		40 (20 + 20)	80 (40 + 40)
Response levels (Doses, LD50 per mouse)	30%, 70% (0.8, 1.3)	616	1232
	16%, 84% (0.6, 1.6)	458	916

Statistical methods for Estimation of LD50

- Various numerical methods depending on
  - assumed model
  - assumed data characteristics
  - actual design
- Analysis of Quantal Response Data, BJT Morgan, Monographs on Statistics and Applied Probability 46, Chapman and Hall 1992.

- There is no magic about numerical methods, and many ways in which they can break down. They are a valuable aid to the interpretation of data, not sausage machines automatically transforming bodies of numbers into packets of scientific fact.
- F. H. C. Marriott
  The Interpretation of Multiple Observations, 1974

Estimation of LD50
Failure of Assumed Model

- There is evidence that the standard deviation of the tolerances of the mice is quite variable on different occasions.
- This suggests that the mice do not come from a homogeneous population.
- Kelly GE (2001). The median lethal dose – design and estimation. The Statistician 50: 41-50

Estimation of LD50: Statistical considerations
Data characteristics: Assumptions vs. Reality

- Mice are randomly assigned to dose levels (treatments)
  - Within 'blocks' if appropriate
  - Randomization procedure must be specified
- Responses of mice (within and between dose levels) are independent

Estimation of LD50
'Randomization' may not be adequate
Data may not be 'independent'

- In an international collaborative study carried out by ten laboratories (Sesardic, Leung, Gaines Das, 2003) there were significant ($p < 0.05$) between cage differences in 25% of LD50 assays.

Data Collection: Records

- Good laboratory practice requires collection of
  - Not only final response (lethality)
  - But also:
    - Animal characteristics, measurements
    - Animal 'scores' at each time of observation, and hence time to final response
    - Experimental details: Caging, Order of Treatment
- Complete records of these data should be available for analysis
  - Opportunities for refinement and reduction, and possible identification of alternate endpoints to lethality

Impact of Experimental Methods (1)

- Experimental methods should be monitored and assessed
- Intraperitoneal injections have been described as 'injections into a black box' (Svendson, 2005, Basic and Clinical Pharmacology and Toxicology 97: 197-199) and where they have been assessed, failure rates of 10% to 20% are widely reported (Claassen V, Neglected Factors in Pharmacology and Neuroscience Research, Elsevier, Amsterdam, 1994)

Impact of Experimental Methods(2)

- Experimental technique has implications for response data and consequently for the statistical model used to describe it and the methods used for analysis
- In an international collaborative study carried out by ten laboratories (Sesardic, Leung, Gaines Das, 2003) there were apparent failures of the larger doses to give the expected maximal response of 100%, and fitted upper limits were commonly about 80%, ranging from 98% to 50% in particular tests

Impact of Experimental Methods (3)
Assume probit model, LD50=1, slope=5

Dose, LD50	Number responding, n=30 per group (assumed injection failures)			
0.6	**5**	5 (0)	2 (3)	5 (0)
0.8	**9**	8 (1)	8 (1)	8 (1)
1.3	**21**	20 (1)	19 (2)	19 (2)
1.6	**25**	23 (2)	23 (2)	21 (4)
LD50	**1.00**	1.06	1.11	1.11
Weight	**1489**	1254	1570	980
Slope	**4.92**	4.46	5.22	3.95
95% Limits	**0.88** **1.13**	0.93 1.22	0.99 1.26	0.96 1.32

Interpretation / Use of LD50

- LD50 estimated for a single experiment is of limited interest
- Implicit in 'precision' of estimate is the intention of extending it more broadly
- LD50 can only be estimated for a single experiment

Interpretation of LD50

- "The dose of botulinum A toxin preparations is expressed in terms of units. The manufacturers state that one unit corresponds to the median lethal dose (LD50) injected intraperitoneally into mice under defined conditions. However, the available preparations are employed at different doses for the same indications, and the units of one preparation cannot be considered to apply to another." (Martindale 32nd Edn)

Interpretation of LD50

- Zbinden and Flury-Roversi stated in 1981 (Arch Toxicol 47:77-99) that '**the LD50 value cannot be regarded as a biological constant**. Through standardization of the test animals and the experimental conditions the variability of the LD50 determinations can be reduced but never fully eliminated.'

Interpretation of LD50

- With ‘cat units, rabbit units, rat units, mouse units, dog units, pigeon units’ bioassay is ‘a subject for amusement or despair, rather than for satisfaction or self-respect’
- And further, use of such units is ‘an insidious means of self-deception’
- Burn JH (1930). The errors of biological assay. Physiological Reviews 10: 146-169

- This leads to the basic principle that biological methods and bioassays are comparative
- Thus ‘while a biological reaction may be used in order to compare the strength of two preparations, one with another, it cannot be used by itself to define the potency of one preparation alone’

Bioassays are comparative

- Concurrent reference preparation and / or controls are essential if bioassay is to be meaningful
- Example: If 50% of mice treated with a dose of toxin die?
  - Negative control and 50% of mice die, deaths may not be due to toxin (contaminated food, bedding?)
- Example: If 0% of mice treated with a dose of toxin die?
  - Positive control and 0% of mice die, conclude mice are not responsive to toxin

Bioassays are comparative

- The European Pharmacopoeia now requires for the assay of botulinum toxin type A that 'a suitable reference preparation is assayed in parallel'.
- Allowance is made for expressing the potency of the product relative to the reference
- As an alternative to this, the reference is required to be within suitable defined limits if the LD50 value produced for the product is to be accepted.

Bioassay

- Unit of measurement: Relative potency
  - Comparative (requiring a reference)
  - 'Unit' of potency defined by reference preparation
- Relative Potency = ratio of doses of reference and test preparations which give the same biological response in the test system
  - To select doses which give the same response is virtually impossible – therefore a range of doses, and dose – response curves

Interpretation of LD50 / Relative Potency

- Estimates from a single experiment are of limited interest
- Implicit in 'precision' of estimates is the intention of extending it more broadly
- Combination of estimates from different experiments

Between laboratory variability for 'absolute' measurement of LD50 and for estimates of relative potency for samples of Pertussis toxin

(From WHO/BS/03.1978; Xing et al. Vaccine 2002)

- Geometric coefficients of variation

–	N	LD50	relative to JNIH-5
– 90/518	6	298%	70%
– JNIH-5	6	314%	-
– Local	4	256%	34%

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biologicals 31 (2003) 265–276
BIOLOGICALS
www.elsevier.com/locate/biologicals

Role for standards in assays of botulinum toxins: international collaborative study of three preparations of botulinum type A toxin

Dorothea Sesardic[a]*, Tong Leung[a], Rose Gaines Das[b]

[a] *Division of Bacteriology, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3QG, UK*
[b] *Informatics Laboratory, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3QG, UK*

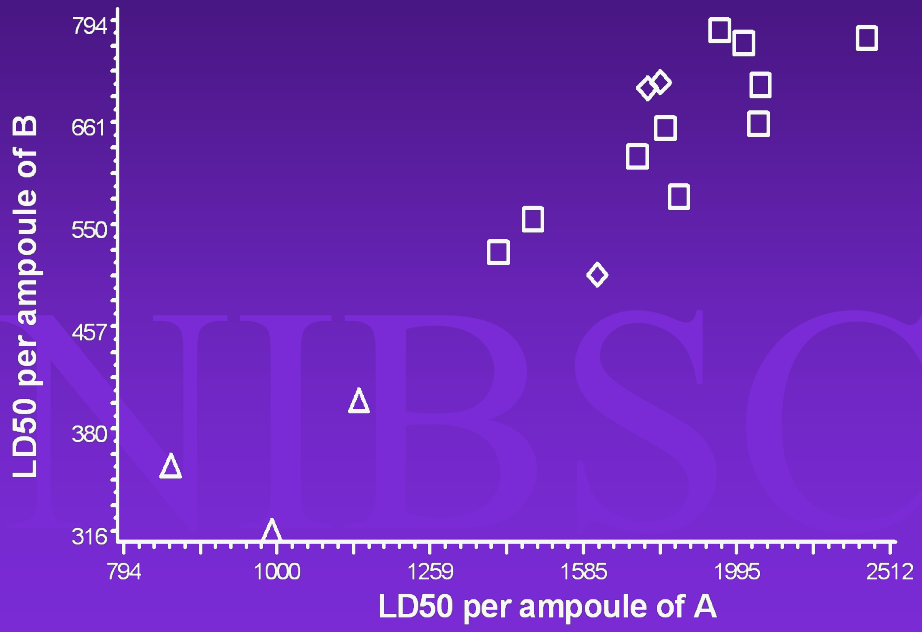
Received 24 March 2003; revised 14 July 2003; accepted 14 August 2003

Abstract

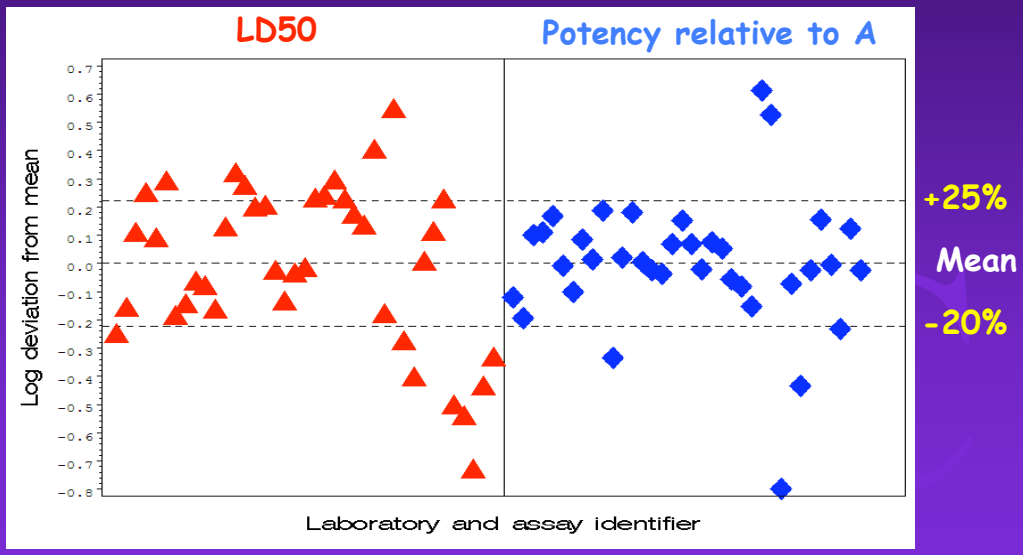
The biological activity of therapeutic preparations of botulinum type A toxin is currently expressed in units defined on the basis of the median lethal intraperitoneal dose of that preparation in mice at 72 h, the LD50 dose. In this study we describe the comparison, by ten laboratories in five countries, of three different formulations of botulinum type A toxin using the mouse lethality

2006 RGD ICCVAM/NICEATM/ECVAM Workshop

Mouse LD_{50} assay is highly variable between laboratories, but laboratory mean estimates are correlated



LD50 and Relative potency estimates (72 hrs) for sample B from all individual assays, (Sesardic et al. 2003)



Relative potency and LD50s using different endpoints

- In the international collaborative study noted above, LD50 values at 72 hrs were about 7% larger than those based on responses at 48 hrs and more than 50% larger than those based on responses at 24 hrs. However, estimates of relative potency do not differ consistently between these times.

Interpretation of relative potency

- Expressed as 'Unit' of activity defined in terms of the activity of the reference standard

- Relative potency reflects relative biological activity as does the LD50, but unlike the LD50 may be system independent to the extent that the biological systems used are specific for the same essential activity

Requirements for the validity of relative potency (comparison) and for reference standards

Functional similarity of the dose – response curves for reference standard and sample is a fundamental condition for assay validity

Reference preparation must be 'appropriate': homogeneous, stable, similar to the preparations it is used to calibrate

Summary
Optimal use of animals for estimates of LD50 / relative potency in individual experiments

- Suitable statistical model
- Optimal design
- Realistic understanding of data characteristics
  - Independence, randomization
- Understanding of experimental methods
  - Interpretation of responses (endpoint definition)
  - Failure rates in treatment
- Complete records of all data, suitably analyzed

Summary
Interpretation of estimates

- Absolute measures such as the LD50 are not biological constants
  - Interpretation is dependent on assay conditions, endpoint
- Comparative nature of biological methods makes suitable controls / reference standards essential
  - Greater independence from assay conditions
  - Improved reproducibility and consistency between and within laboratories
  - Wider opportunity for development / use of alternate endpoints to lethality and for refinement and replacement

2006 RGD ICCVAM/NICEATM/ECVAM Workshop