# Statistical aspects of the design, analysis and interpretation of chronic rodent carcinogenicity studies of pharmaceuticals

## Some comments on the FDA draft guidance document

Author : P N Lee
Date : 8.8.2001

### Background

I was a co-author of the 1980 guidelines by Peto et al[1] and the 1986 IARC Monograph by Gart et al[2] on the design and analysis of long-term animal experiments. I have been operating as an independent statistical consultant since 1979 and have analyzed numerous long-term animal studies for various clients, many of which have been submitted to FDA. My company, P.N. Lee Statistics and Computing Ltd., has developed the ROELEE program which allows data entry, reporting and analysis of long-term animal studies and which is used by various companies, contract houses and independent pathologists and by ourselves for our consultancy work.

I have recently become aware of the draft guidance document and hope that my comments are not too late.

### General considerations

The draft guidance document is very clear and what is said is generally very appropriate. However there are a number of extra points that could very usefully be made and a few other issues, which I elaborate on below.

### Analysis of pre-neoplastic and non-neoplastic lesions

In carcinogenicity studies, pathologists typically report not only the presence of tumours, but also the presence of pre-neoplastic lesions (such as focal hyperplasia) and of non-neoplastic lesions. The guidance document (except in one very brief reference on page 27) refers only to the analysis of tumours. This is unfortunate for two reasons.

First, I believe it is important to analyze pre-neoplastic lesions to assess carcinogenicity properly. The interpretation of a marginally significant dose-related increase in tumour incidence may well be crucially affected by whether or not there is a similar relationship of dose to the corresponding focal hyperplasia.

Second, it would be useful to make the point that the need for age-adjustment applies equally well whether one is analysing tumours, pre-neoplastic lesions or non-neoplastic lesions (as is the need to define whether a lesion is fatal or incidental). Some laboratories with which I have contact use unadjusted chisquared or Fisher exact tests for analysis of non-neoplastic lesions, which is total nonsense when there are survival differences. One little sentence by the FDA could have a great educative effect.

Combination of tumours

There is no mention of whether or when one should carry out combined analysis of different tumours (e.g. benign and malignant tumours of the same site, leukemias presenting in numerous tissues) or of tumours and related focal hyperplasia. Clearly some combinations are essential, others are meaningless. A suitable reference to McConnell et al (1986)[3] would be useful. I note on page 35 that in Table 15 you give an example where analyses are carried out of hepatocellular adenoma and carcinoma separately but not combined. One hardly wants to encourage people to think they can get away with omitting essential combined analyses.

Recording of whether a tumour is fatal or incidental

In the general situation, tumours are only seen for the first time at post-mortem and there is only a single terminal kill at the end. In this situation, in order to carry out proper analysis, it is necessary either to decide whether or not the tumour killed the animal (and use the Peto method) or to estimate when the tumour arose (and use the same method as used for observable tumours). The latter is not generally done, presumably as it involves too much guesswork, but the former is.

There are four issues I have with the draft guidelines. The first is a semantic one and concerns use of the word "lethality". To my mind a tumour is lethal if it has the potential to kill,

and it is quite possible for an animal to die for some totally unrelated reason with a tumour that would have killed it shortly afterwards.

The second issue is that some of the discussion on what is required for an analysis is rather confused. The guidelines appear to state that one needs information on cause of death and lethality to carry out the Peto analysis. This is not so. The Peto method only requires one to know whether a particular tumour in a particular animal was or was not responsible (directly or indirectly) for the death, i.e. whether it was the cause of death. Whether the tumour was slow growing and harmless and would never have killed the animal (totally non-lethal) or was aggressive and would very soon have killed the animal had it not died of something else (lethal) is irrelevant; it would still go down as an incidental tumour. The discussion relating to Haseman (1999) in the first paragraph of page 9 is off the point - with regard to classifying a tumour as fatal or incidental there is no "continuum". A tumour either killed the animal or it did not. If there is doubt, one can use the probably or definitely fatal or incidental classifications suggested by Peto et al (1980)[1] and run a sensitivity analysis. For fatal tumours the real issue is whether the relative numbers alive in each group at the time the animal died are a good approximation to the relative numbers tumour-free in each group at the time of tumour onset. If the time from onset to death is short, it clearly will be, but if not, it may not be. There is an excellent discussion in Gart et al (1986)[2] on pages 13-15 on the potential for bias in the analysis of fatal and incidental tumours which might well be referred to.

The third issue I have with the draft guidelines in this area is that I am unhappy that they are not critical enough of carcinogenicity studies that do not provide information on cause of death. They should be, as without such information it can be impossible to analyze the data properly.

Finally, and related to the third point, the guidelines should make it clear that though the poly-k and ratio trend tests might perform well under certain assumptions, one can easily construct situations where they come up with totally the wrong answer. For example, consider a simple situation where in group A all the animals get a tumour at week 50 which would kill them at week 80, while in group B all the animals get a tumour at week 70 which would kill them

at week 100. In group A there is no mortality from other causes, so all animals die at week 80 with the tumour. In group B there is severe mortality from other causes, killing them all at week 80. Any poly-k type test would estimate that there was no treatment difference between A and B in relation to the tumour of interest, as the data for the two groups are the same. The Peto test would find a clearly higher carcinogenicity in group A, which is correct. Indeed, if one assumed the mortality from other causes killed all the animals in group B at week 75, the poly-k test would come up with an answer in the wrong direction! Even disregarding the unsatisfactory arbitrariness of whether one uses $k = 3$ or some other value in a poly-k test, the apparent endorsement of this test by the FDA is extremely worrying when the test has the propensity to produce nonsense results.

Treatment of missing data in the Peto method

The guideline should give help about what to do with missing data. A standard situation is that, for an animal, one knows when it died but a section of a particular tissue is lost or too autolysed for examination. When analysing tumours of that tissue it should be made clear that it is incorrect to remove that animal totally from the analysis. As one knows that the animal did not die from the tumour before the week of death, it should be included in the at-risk population for fatal tumour analysis for each week before its death, though omitted from the at-risk population from fatal analysis for the week of its death, and from incidental analysis for the period of its death.

Role of the procedure for histopathological examination

In my experience, it is really quite often the case that one has to analyze tumours for three different types of tissues:

(a) Those scheduled for routine microscopic examination in all groups,

(b) Those scheduled for microscopic examination only if abnormal at post-mortem,

(c) Those scheduled for microscopic examination except in low and middle dose terminally killed animals with no abnormality at post-mortem.

While the analysis for tissues of type (a) is straightforward, subject to the provision noted for missing data, there are issues to be raised about tissues of type (b) and (c) which should

certainly be discussed in the guidelines. My own procedure is as follows.

For tissues of type (b) I usually consider all animals at risk, whether or not a section was examined. One is then analysing the proportion of all animals with a tumour large enough to be detected at post-mortem. Omitting from analysis animals that are normal at post-mortem is not appropriate, as the tumour is often the reason for the abnormality. If 2 control animals and 20 dosed animals are found to be abnormal at post-mortem and all have a tumour, a comparison of 20/50 vs 2/50 seems more appropriate than one of 20/20 vs 2/2. However, there is the possibility that treatment might cause some other condition resulting in post-mortem abnormality, thus enhancing the chance of a small tumour being observed microscopically in that group. There is a case for limiting attention only to conditions that would have been seen at post-mortem.

For tissues of type (c), where there is a different examination procedure in the different groups, there is an obvious danger of bias in the analysis. One can avoid bias by excluding from analysis animals in the low and middle dose groups when they reach terminal kill (though, as noted above, one must include them in the at-risk population in previous weeks). However, in some studies the proportion of animals reaching terminal kill is large and the loss of information from omitting numerous tumours seen in low and middle dose terminally killed animals from analysis can therefore also be large. Here it is often worth carrying out two additional analyses to gain further insight. In the first additional analysis, only animals that were examined microscopically are considered at risk, while in the second additional analysis all animals are considered at risk. The first analysis tends to bias risk estimates upwards in the low and middle dose groups, but any significant decreases should be real. The second analysis tends to bias risk estimates downwards in these groups, but any significant increases can be taken as real.

Clearly there is a need for this general issue to be discussed as there is a huge potential for incorrect analyses. There is a mention on page 28 that some sponsors choose not to do histopathologic evaluation of all treatment groups, but this is rare. It is much commoner in my experience to have a situation like (c) above.

Dose metameter

In the trend analyses, D is defined on page 11 as actual dose. While I believe this is the correct procedure generally, the guidance document would do well to comment somewhere on the validity of the commonly used alternative procedure of using dose metameters of 1, 2, 3, 4 for control, low, middle and high.

Fatal analysis first

On page 10 it is correctly noted that $R_k$ for incidental tumours is the number of animals that have not died of the tumour type of interest but come to autopsy in time interval k. However, the point is easy to miss that the logical way to do Peto analysis is to do the fatal analysis first, then remove the fatal cases, then do incidental analysis. It might be better to make this clearer and indeed to have the section on analysis of fatal tumours before that of incidental tumours.

Use of the term "mortality-independent"

Although I was a co-author of the original Peto monograph that uses the term "mortality-independent" I have never liked it as being just too obscure. I would much prefer the guidance document to use the term "observable" (or perhaps "observable in-life") as the routine and attempt to get away from the use of "mortality-independent" (though one could in one place refer to this as the term used by Peto et al).

Pairwise tests

There is no information in the guidance as to how to do them. Should one use the trend formula with only two groups, or apply a standard continuity correction? The latter I would have thought.

Tails for p-values

I believe it is standard practice to assume that all p-values presented in a scientific paper are two-sided unless it is specifically stated that they are one-sided. Though in some places, e.g. the bottom of page 14, it is specifically stated that one-sided p-values are one-sided, there are plenty of cases where what are actually one-sided p-values are not described as such. For example, Table 6 should make it clear that the p-values are one-sided and the text on pages 27

and 28 could easily be misread without clarification.

Survival analysis

On page 8, when discussing tests for survival differences, why is the possibility of using a Peto fatal analysis not included? If you had a study where all deaths were because of one specific tumour, analyses of survival and of tumour incidence should be the same.

Minimal points

On page 4, section IV A, para 1, line 8, Peto and McKnight are both misspelt. McKnight has also lost her "n" further down on page 4 and near the end of page 5.

On page 10 on the first line, "notation" should be singular, and the same applies to the heading of Table 3 on page 12.

On page 14, the third time interval in Table 5 has lost its name "81-106". Also, the footnote to that table is incorrect - it should read "the expected tumor prevalences were calculated under the null hypothesis that within each time interval prevalence is independent of treatment". There are an infinite number of null hypotheses that produce no trend, but in only one is the incidence equal in each group! Note I have changed "incidence" to "prevalence". This change should also be made in three places in the paragraph preceding the table.

On page 15, in the definition of $R_k$, "at risk" should be defined.

At the bottom of page 16 "prespecified" is misspelt.

On page 22, in the first line below Table 12, "trend" has lost its "r".

On page 24 in the last paragraph and twice on page 25 in the first paragraph, "Bieler-Williams" has lost its "s".

On page 28, in paragraph 3, line 3, "Lagakos" is misspelt.

## References

1. Peto R, Pike MC, Day NE, Gray RG, Lee PN, Parish S, *et al.* Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. In: *Long-term and short-term screening assays for carcinogens: a critical appraisal,* Volume Supplement 2. Lyon, France: IARC, 1980;311-426. (IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans.)

2. Gart JJ, Krewski D, Lee PN, Tarone RE, Wahrendorf J. *The design and analysis of long-term animal experiments,* Volume 3. Lyon: International Agency for Research on Cancer; 1986. (Statistical methods in cancer research.) IARC Scientific Publications No. 79.

3. McConnell EE, Solleveld HA, Swenberg JA, Boorman GA. Guidelines for combining neoplasms for evaluation of rodent carcinogenesis studies. *J Natl Cancer Inst* 1986;**76** :283-9.