Huntingdon
Life Sciences

Dockets Management Branch (HFA-305)
Food and Drug Administration
12420 Parklawn Dr., rm 1-23
Rockville
MD 20857, USA

7 August 2001

Dear Sir/Madam,

Guidance for Industry: Statistical Aspects of the Design, Analysis and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals (815.dft).

[pursuant to the notice issued by the Food and Drug Administration in the Federal Register of 8 May 2001 (Volume 66, number 89) reference 23266-23267.]

We have reviewed the above document and find it well-presented and very interesting. We would like to make some comments on the content and also point out a few minor errors. Please therefore find our response attached.

This represents a joint response of the Statistics Departments of Covance Laboratories Europe and Huntingdon Life Sciences (HLS). These two organisations are CROs with long experience of this area, having conducted, analyzed and reported hundreds of carcinogenicity studies over the last thirty years.

Yours sincerely,

Graham Healey, Head of Statistics, HLS
Christine Gardner, Head of Statistics, Covance Laboratories Europe
Ann Gradwell, Principal Statistician, Covance Laboratories Europe
Richard Brammer, Senior Statistician, HLS

cc: CG/AG (Covance), RB, S.McCormick (cover only)

01D-0194

I
C10

Dept. of Statistics, Huntingdon Life Sciences, Woolley Road, Alconbury, Cambs., PE28 4HS, UK Phone (44) (0)1480-892112

For the attention of:

Dockets Management Branch (HFA-305)
Food and Drug Administration
12420 Parklawn Dr., rm 1-23
Rockville, MD 20857, USA

Guidance for Industry: Statistical Aspects of the Design, Analysis and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals (815.dft).

[pursuant to the notice issued by the Food and Drug Administration in the Federal Register of 8 May 2001 (Volume 66, number 89) reference 23266-23267.]

This document represents the joint response of the Statistics Departments of Covance Laboratories Europe and Huntingdon Life Sciences (HLS).

Section and line numbers correspond to those of the Guidance for Industry draft document.

Authors:

Graham Healey, Head of Statistics, HLS
Christine Gardner, Head of Statistics, Covance Laboratories Europe
Ann Gradwell, Principal Statistician, Covance Laboratories Europe
Richard Brammer, Senior Statistician, HLS

7 August 2001

SECTION III p2: Validity of Design

Lines 74-77: Randomization

- Given the importance of randomization for valid statistical inference, this section should really be expanded. There are many levels at which clear advice could be offered. For example:

  Allocating animals to cages (eg gang-housing)
  Allocation of treatments to animals or cages (eg by bodyweight).
  Allocation of cages to batteries or racks (eg latin squares).
  Rotation of batteries.
  Dosing order.
  Order of in-study manipulations.
  Order of sacrifice and autopsy.
  Order of slide-reading.
  Deployment of multiple pathologists.

Lines 79-92: Slide-reading

- We add that if there was a clear separation between control and treated group lesions then "open-reading" would be fine. This is not always the case, however.

Lines 94-100: Interim sacrifice

- It might be mentioned here that from a purely statistical point of view, interim sacrifice generally increases the number of animals required but provides only a limited amount of statistical information as far as tumour incidence is concerned.

SECTION IV p4: Statistical Methods

SECTION IV.A p4: Overview

Separate sexes

- It maybe worthwhile stating here that the data from the two sexes are almost always analysed separately. Combined analysis is possible however.

General approach

- We agree with the authors that the Peto methodology as described in the document is currently the best method for the analysis of tumour data.

Low tumour levels

- We believe it to be a useful practical rule to apply statistical analysis only to those tumour types with a total incidence over all treated groups of at least a certain number, say two or three. No statistical analysis is performed if the total incidence in the treated groups is less than this. It is a practical measure reflecting the fact that tumours occurring no more than once cannot be found to have a statistically significant increased incidence.

SECTION IV.B p6: Adjustment for Mortality

Lines 292-298: Survival analysis.

- In view of recent documents coming from the FDA, it should be clarified how many types of analysis are required. To maintain good statistical practice only one method should be presented instead of a battery of tests (Cox, Wilcoxon, logrank, etc) as suggested.

SECTION IV.C p8: Analysis without cause of death information

SECTION IV.C.1 p8: Context of observation

Line 311:

- The sentence starting "Tumours that are..." is ambiguous. It should read something like "neither directly nor indirectly". (See PETO *et al* p328).

Lines 331-341:

- The end of this section is ambiguous. Is it saying "100% of tumours of a certain type are fatal or 100% are incidental" or is it "each tumour is either 100% fatal or 100% incidental" ?

SECTION IV.C.2 p9: Analysis of incidental tumours

Line 411:

- There is a mistake in this line [$E_i=\sum E_{ik}$]

Table 3:

- We note that by the definition of Pik the quantities P.k are all equal to 1.

Lines 415-430: Hypothesis test

- There is no mention of continuity corrections here (nor in fact anywhere in the document), yet the literature constantly refers to them.

- It should be made clear that one-tailed tests are being used here.

- There is no discussion of the use of nominal versus ordinal dose levels. There can sometimes be a big difference between them. Our preference is for nominal dose levels.

- One further possibility is to apply a "non-linearity" test (eg a chi-squared test with k-2 degrees of freedom) for the deviation from the linear trend. If the test was significant, then the pairwise results might be preferred to those for the trend tests.

Lines 441-447: Methods of determining intervals

- A sensible recommendation (GART *et al*) is to combine intervals if, for one interval, there are only deaths in one group.

SECTION IV.C.5 p16: Analysis of mortality-independent tumours

- A palpable tumour can be found in an incidental context (the animal dying before the tumour was palpated). The text should clarify how all data for a tumour-type occurring in several types of classification (incidental, fatal or palpable) in the same study can be included in the same analysis. The easiest approach is to simply include all the different types of strata in the same test.

SECTION IV.C.6 p17: Exact tests

Lines 599-603: Asymptotic vs Exact

- It is common practice to have a cut-off such as "When the total incidence for a particular tumour is greater than or equal to 2, but less than 10, exact tests will be applied." Is it acceptable to use permutation tests at all levels of tumour incidences ?

Lines 611-612: Fatal tumours

- The limitations of applying exact tests to fatal tumours, where an individual animal may contribute to multiple strata, should be expanded upon. Is the assumption of independent strata adequate ?

SECTION IV.D p21: Analysis without cause of death information

Line 877-878:

- The wording is unclear. Should the phrase "results only in a" be "requires only one" ?

SECTION IV.E p25: Dual controls

Line 942:

- It is not clear what the "first" and "second" cases are.

Lines 949-951:

- There are several possible approaches:

First approach:	"Significant if both T>C1 and T>C2".
Second approach:	"Significant if either T>C1 and T>C2".
Combined:	"Significant if T>C1+C2".

We agree entirely (we have performed simulations) that "the first approach" is conservative, and the "second approach" liberal. The combined analysis turns out about right. Hence we see no reason to split the controls at the analysis stage. Analysis of historical data, as reported in the literature, suggests that differences between identical control groups are purely random (ie no extra-binomial variation).

The guideline recommends checking to see whether the control groups are similar, combining them if they are and doing both combined and separate analyses if not, looking at historical control data to decide how to interpret the separate analyses. This guideline is rather vague and time-consuming. It would be simpler and statistically appropriate to always just do the combined analysis.

SECTION V p26: Interpretation

Line 967:

- We do not necessarily know at the time of analysis what other studies are being or will be performed. Therefore at the statistical analysis stage it is generally not feasible to consider "programme-wise" error rates, only "study-wise".

SECTION V.A p27: Adjusting for multiple tests.

- We agree that this section refers more to "interpretation" than to "analysis". Therefore we consider it essential to present the original unadjusted *p*-values first.

  In particular, the switch from using a 0.05 level to a 0.005 level is very major.

- We note that only false-positive rates were considered. We believe the absence of discussion of false-negative rates in a statistical context (as opposed to Section V.B) to be a quite serious omission.

  For example, if, for a common tumour, $p \approx 0.05$ for both sexes in both species, this would be substantial evidence of a carcinogenic effect but would not be "significant" by these rules. Analysis of combined sexes/species (as recommended by GART *et al* if carcinogenic effect is similar) in addition to the separate analyses would be useful in these circumstances.

  Again, using $p<0.01$, to obtain 80% power for a tumour of (just above) 1% control rate would require up to 20% incidence in the treated group, indicating very low power.

- We note from simulations that if there is to be a common/rare threshold, then given the distribution of control group tumour rates and the false-positive rates for low frequencies, then 3% seems to be about the best. A level of 1% is certainly too low, and it also suffers from there being too many tumours around 1%.

- Correlation between the tumour-types would considerably affect the false-positive rates. This correlation is currently unknown.

Lines 1045-1049: Incomplete examination of tissues

- If only the control and high dose groups are fully examined, then tests should be performed on stratified incidence including only the pooled control group and the high dose group.

- If none of the groups have all animals examined, then a full statistical analysis cannot be performed.

SECTION V.C p30: Historical control data.

Lines 1112-1114: Database

- We strongly support the statements concerning the database quality and study design factors (eg diet, pathologist, amount of tissue examined, etc).

Line 1140-1142:

- The wording here is unclear – the sentence seems to have been split.

Line 1142:

- We agree that the range is not ideal, and we believe that percentiles or confidence intervals are preferable. We note that the table still shows ranges.
- We note that only a few Sponsors and CROs will have adequate historical data.
- The document rightly suggests that comparisons will be difficult if there are survival differences between studies (SEEWALD).

SECTION VI p32: Presentation

Line 1220:

- There are no suitable descriptive statistics for tumours. Please clarify.

Table 14, p34

- We note that this table is not normally produced at the statistical analysis stage or by statisticians.

Table 15, p35

- This is not a very well-formatted table. In particular we believe there should be no need to present both asymptotic and exact *p*-values. Please clarify whether this is just one possible format or if it pretty much represents the required format.

Table 16, p37

- We note that no overall point estimate is presented. Also, we agree with the text (line 1142) when it suggests that ranges are not really appropriate, but the table does not reflect this view.

Data submission

- There is an urgent need to clarify the various recent documents in circulation (1997, 1998, 1999) concerning the format of data submission. We understand that the 1999 format is the only one acceptable to all divisions of CDER.

ADDITIONAL TOPIC not mentioned in the Guidance document

Amalgamation of tumour types

Some discussion of this topic could be included.

- For certain tumour-types it does not make sense to analyse benign tumours alone.
- Hyperplasia or other non-neoplastic lesions may or may not be included. A categorical classification of tumours into None/Hyperplasia/Benign/Malignant could be attempted. It has been noted (GART *et al*) that it is dubious to analyse non-neoplastic lesions separately, if the

Line 1142:

- We agree that the range is not ideal, and we believe that percentiles or confidence intervals are preferable. We note that the table still shows ranges.
- We note that only a few Sponsors and CROs will have adequate historical data.
- The document rightly suggests that comparisons will be difficult if there are survival differences between studies (SEEWALD).

SECTION VI p32: Presentation

Line 1220:

- There are no suitable descriptive statistics for tumours. Please clarify.

Table 14, p34

- We note that this table is not normally produced at the statistical analysis stage or by statisticians.

Table 15, p35

- This is not a very well-formatted table. In particular we believe there should be no need to present both asymptotic and exact *p*-values. Please clarify whether this is just one possible format or if it pretty much represents the required format.

Table 16, p37

- We note that no overall point estimate is presented. Also, we agree with the text (line 1142) when it suggests that ranges are not really appropriate, but the table does not reflect this view.

Data submission

- There is an urgent need to clarify the various recent documents in circulation (1997, 1998, 1999) concerning the format of data submission. We understand that the 1999 format is the only one acceptable to all divisions of CDER.

ADDITIONAL TOPIC not mentioned in the Guidance document

Amalgamation of tumour types

Some discussion of this topic could be included.

- For certain tumour-types it does not make sense to analyse benign tumours alone.
- Hyperplasia or other non-neoplastic lesions may or may not be included. A categorical classification of tumours into None/Hyperplasia/Benign/Malignant could be attempted. It has been noted (GART *et al*) that it is dubious to analyse non-neoplastic lesions separately, if the

lesion is an intermediate step to a tumour. In this case, it is sensible to analyse tumours and non-neoplastic lesions combined.

- If the tumour were an amalgamation of two benign or two malignant tumours, then if an animal had a non-incidental and an incidental tumour, only the non-incidental tumour would be included in the analysis.

REFERENCES

GART, J.J. *et al.* (1986) *Statistical Methods in Cancer Research. Volume III - The design and analysis of long-term animal experiments. IARC Scientific Publications No.79.* International Agency for Research on Cancer, Lyon.

PETO, R. *et al.* (1980) 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. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Supplement 2,* pp.311-426. WHO International Agency for Research on Cancer, Lyon.

SEEWALD, W.(1994) Time Trend in Historical Controls for Tumour Incidences in Long-term Animal Studies. *Applied Statistics* **43**(1): 127-137.

PRIORITAIRE

En cas de non-remise prière de retourner à postach 2012 22764 HAMBURG ALLEMAGNE

Deutsche Post

port payé 60514 frankfurt ALLEMAGNE

Brief / Infopost

Dockets Management Branch (HFA-305)
Food and Drug Administration
12420 Parklawn Dr., rm 1-23
Rockville
MD 20857
USA