

Rebuttal – Leonard report October 2006

1.1 – Terms of the review

This person was asked to review the statistical analysis of the ESR report. The study author would like to know if there are any relevant qualifications of this individual which would lend credence to his critique considering the combined experience of the team of researchers who carried out the study, and the long list of internationally recognised experts who gave reviews of the study and report.

2.1 Mr Leonard claims that the sample sizes are too small to reach the main conclusions in the report.

Before conducting the study we undertook what is called a “power calculation” which tells us approximately how many people we would need to sample in order to see a significant difference between two groups if a difference does in fact exist of a certain magnitude. At the time the only data available to use as a basis for this were the estimated releases of TCDD from the 1986 bursting disc failure, and the 1997 MfE serum dioxin study, which informed us to some extent about the population variability in baseline serum TCDD.

The main concern about small sample sizes is that they increase your likelihood for generating false positive or false negative results. However, if you end up seeing a difference of much greater magnitude than you expected to find, and you see this across all age/gender groups, then your need for a large sample size diminishes. There is simply no way the results we found could have occurred by chance. In other words if the evidence is strong enough, the proof is then in the pudding and you don’t need to throw additional samples into the analysis to answer your central question.

Mr Leonard assumes that there was a much larger pool of highly exposed individuals whom we did not include in the sampling, thus compromising the study statistics, biasing results downward, and giving artificially large estimates of variability. Quite to the contrary, we successfully sampled the vast majority of those individuals expected to have significantly elevated TCDD levels. Those who showed the very highest levels were elderly people whom had lived in the area for the longest period of time. We specifically reported the mean values and range for this group to ensure that the highest exposed subgroup would feature statistically in the report.

Some of the people we wanted to sample could not give blood due to health reasons, and while this is unfortunate, it was an inescapable reality and no fault of the study design. Adding people who were not expected to have highly elevated TCDD levels into the sampling groups would have had the effect of diluting the magnitude of the exposure in the group. This would have clouded the exposure assessment and would have done the highly exposed individuals a disservice.

In summary, it is an oversimplification to state that more samples would have made a better study.

In our case we had several core questions we were asked to address:

1. Is there evidence of dioxin exposure in the Paritutu community?
2. If so, what was its source (incinerator vs other sources)?
3. If so, what was the geographical extent of the exposure?
4. If so, what was the magnitude of the exposure?
5. If possible (and this was added in at the very end), can we rule any particular timepoints in or out as periods of likely exposure?

We had originally proposed to carry out all the sampling at once. The Ministry was opposed to this, and after consultation with its advisory committee (OTAG) decided we should instead carry out collecting a few samples first to see if there was evidence of exposure or not (argument being that if we saw absolutely nothing in the first lot of highest modelled results, there would be little point in wasting tax payers' money measuring all the rest of the samples).

But in fact after the first 24 samples we had already answered the first three principal questions, and there was, beyond a shadow of a doubt, evidence of TCDD exposure in the community. For the fourth question, we had an estimate of the magnitude of exposure, but since people selected for the study were only those estimated to have the highest TCDD exposures (of those who came forward to us), and because many of these people were elderly (this posed additional scientific questions as well, like what is their TCDD elimination half-life, because this is simply not known for people over 70), it was not a representative cross-section. And for the fifth question, we lacked people in the first 24 who lived in the exposure zone between certain time periods (by pure chance – not by design). In particular we wanted to try and isolate time periods that would tell us about the significance of the 1986 explosion, the 1972 explosion, and the onset of the Clean Air Act in 1974. However, this was the most difficult question to answer, and we make this clear in our discussion section about the uncertainty with regard to the evidence as it relates to timing of exposure.

In fact we state in the report that temporal variations in exposure remain an area unaddressed by the current study.

We told the Ministry that this was going to be a difficult last question to answer and they insisted that we continue to sample additional people to try and seek more clarity about questions 4 and 5. There was also a contractual agreement to carry out a minimum of 50 samples. So for these reasons we carried out the second sampling group.

2.2 The two sampling groups cannot be combined

Now, if after all the samples were analysed for differences across times of residence and a clear, “black and white”, exposure window was identified, then it would have been essential to segregate the overall sample group accordingly (e.g. pre-1973 vs post, etc). But this was not the case. There was instead evidence of exposure in all the time periods between 1962 and 1987. We also did not have the luxury of picking and choosing ideal candidates from all time periods, but were rather faced with what was available. Therefore the distinction between the first and second lots of samples was artificial and there was no justification for keeping their results separate. People in these two lots were in fact neighbours with overlapping residence

times. The combined analysis allowed us to increase our number of demographic groups and better characterise a cross section of the community. And while it is true that more people in the second lot lived in the area for shorter durations, we made sure this didn't "wash out" the estimates of body burden for those most highly exposed by reporting the statistics on the long term residents separately. There you get a much better feeling for the potential magnitude of exposure.

2.3 The report does not contain details of the time period analysis

True. We did not think it would be informative to add this analysis to the report, since it was mostly a semi-quantitative result, and that is why we clearly state in the discussion on page iii that the temporal variations in exposure between 1962 and 1987 remain unaddressed in the current study. Our estimated back-calculations to peak body burden are simply to give the reader an approximation of the magnitude. Further, more detailed examinations of the existing data could shed additional light on more precise possible peak body burdens, but would not be expected to change the estimate to a degree that would be meaningful in terms of health risk.

2.4 There is an error in this sentence in the final report. This is the only substantive error in the comments provided by the reviewer.

20.1. In the 2005 Report (Pg18) states that of the 37 people who lived in the area less than 15 years, only one was demonstrably elevated (17.9pg/g). The next highest was 7.1pg/g.

It should read "...of the 38 people who lived in the area less than 15 years, two were demonstrably elevated (17.9 and 14 pg/g)" This has the effect of changing the paragraph on page 18 of the report to "...people who lived in the area for less than 15 years had a mean TCDD level of 3.6 ppt, and those who lived in the area for more than 15 years had a mean TCDD level of 14.7 ppt." It's very helpful that the reviewer picked up this mistake, and an erratum has been prepared. The lead author apologises for any confusion it may have caused. However, it makes absolutely no difference in the final analysis or any of the conclusions. The figure 4 in the report remains exactly as it currently is.

2.5 There is a claim that the report contains contradictions or inconsistencies of consequence. The example below is given by the reviewer, but I fail to see how these two statements are conflicting. The reviewer has not indicated how they can be interpreted differently, or what consequences result from this, so I am unable to address the concern

- 11.7. There is no reference in the body of the report to the effects of kai moana. The reader is reliant on the statement in the summary I reiterate

“There is no indication of a significant contribution to exposure from root vegetables, protected fruit, (Citrus, feijoas, etc) poultry or eggs or kai moana (seafood)”

- 11.8. The ESR annual report 2005 under the heading “Dioxin Study – Paritutu – Key Points” stated:

“No link was found between dioxin levels and eating homegrown root vegetables protected fruit such as citrus, poultry/eggs or local kaimoana”

- 11.9. The ESR Report can be interpreted differently to the summary in the 2005 Report.

2.6 The historical body burden calculation is incorrect.

There is no way to know with precision what the peak TCDD serum levels were in Paritutu. Given the numerous uncertainties that exist and are discussed in the discussion section of the report, one can calculate any of a wide number of scenarios which give different historical body burden levels. We presented several possible back calculations using specified assumptions to give the reader a feel for the magnitude of the historical levels. We believe that the magnitude is broadly comparable to that of the Seveso Italy Zone B cohort, though could be slightly higher on average than this group (as shown in the report). The reviewer has ignored the discussion of uncertainties in the report and presumed that a high degree of certainty exists about such a calculation. This illustrates his lack of understanding about the nature of the results, and the inherent variability and unknowns surrounding recreation of historical body burdens.

In our calculations, we did not merely take TCDD values and calculate backward to 1987 for all participants – instead we took the last residence year for each individual, or the year 1987, whichever was earlier, and calculated individual historical levels using a published and conservative 7.1-year half life. We did this because we saw evidence of TCDD exposure even in later residence times (approaching 1987). So our assumption was that an individual’s last exposure would have been the last year they lived in the exposure area (1962-1987), or else 1987, whichever was the earlier year. This is all explained in the exposure reconstruction, together with the appropriate uncertainties. If one reads this section of the report, it can be seen that we did not intend to present these exposure reconstructions as authoritative final analyses, but as our best reasoned estimates. We also showed a possible historical body burden for the highest exposed individual, which gave a value of 225 ppt. This was to help characterise the potential range of exposures and body burdens. The reviewer’s calculations of 112 and 93 ppt are entirely consistent with this highly variable and uncertain calculation. Mr Leonard decries the variabilities involved in the study yet seems to find that a difference of 93 and 98 amounts to something significant (even

considering this difference to indicate some sort of error). This is a natural consequence of enlisting a reviewer, whom while is no doubt very competent in his particular field, has no relevant qualifications for such a review.

2.7 The report and raw data should be reviewed by a biostatistician

The report was reviewed (even co-authored) by a biostatistician and several PhD scientists with extensive experience in environmental exposure and risk assessment. A second PhD biostatistician performed the power calculation already described earlier.

Personal note from the Study Author:

This study, from its consultation phase through to its protocol and design to its final report, was reviewed by an international group of leading scientists, and was found to be well conducted, using sound methodology and having sound conclusions. The work has been presented at two international fora: GeoHealth (Wellington, 2004), and the International Society for Environmental Epidemiology (Paris, 2006). A manuscript is in preparation and will be submitted to a major international journal in 2007. The study found clear evidence of TCDD exposure in the Paritutu community, dating from the 1962-1987 245-T production period. We successfully characterised the presence and magnitude of this exposure, its source (fugitive emissions during production and not incineration), its geographic boundaries, ruled out some historical events as being major contributors to the exposure (the 1986 explosion) and qualitatively assessed the timeframe of the exposure.

The review gave the study authors zero credit for successfully addressing any of these issues. Additionally, the reviewer never mentions any of the caveats presented in the discussion or the discussion of uncertainties in the report. The review was therefore unbalanced, misrepresented the study findings, and was not intended to inject constructive criticism into the debate and/or provide a mechanism to open reasoned discussion or to advance new hypotheses worthy of exploration (any of these courses of action would have been very welcome by the research team). The intent seems instead to have been to discredit the study from start to finish, to cast as much doubt and uncertainty about the indisputable findings as possible, to impugn the research team and institutions involved, and to generally provide the television network with a 'government cock-up, cover-up' story. This is unconscionable in this author's view, from the standpoint of decent and fair journalism, and from the lack of professional ethics of the reviewer in providing this review to the TV journalist without a word of contact or earnest effort to discuss areas of concern with ESR or any of the study authors beforehand.

I have acknowledged the presence of two errors in the report, both inconsequential to the main study conclusions, and of no impact to any of the study report tables or figures. Yet there has been no small effort to inflate, to a dramatic level, the significance of these minor errors both in the review and in the television media. The error in the Appendix was no more than typographical and is virtually meaningless in the context of the report which used all the correct values. The other error is also minor, and when corrected changes none of the study report's main conclusions. An

erratum involving a change to the text is warranted for these two changes, and this has been drafted and sent to ESR and the Ministry of Health.

I question the motives of the television network and the reviewer in spending 18 months quietly searching for flaws or fomenting areas of doubt about what is a very well-reviewed and accepted study, and above all in exploiting community fears and distrust in government in order to gain media attention.

Having personally met a number of the Paritutu study participants, I am saddened that the report had any flaws whatsoever in it, and I apologise to them for that, but I hope sincerely that they realise that this politically motivated and unscientific review by someone without relevant qualifications, makes absolutely no change to the conclusions of the report.