Critique of the validation studies conducted to date of *in vitro* methods for determination of leaching rates of phthalates from PVC toys (conducted by TNO and LGC), and of the *in vivo* study underlying the validation of the Dutch methodology (as conducted by RIVM)

submitted by Greenpeace International

for consideration by the European Community Scientific Committee on Toxicity, Ecotoxicity and the Environment (CSTEE), September 1999.

Greenpeace Research Laboratories Technical Note 03/99: David Santillo[1], Paul Johnston[1] & Axel Singhofen[2]

Summary

Critiques of each of the above mentioned studies are presented in the following three papers. A major concern, besides those fundamental concerns relating to the overall approach of using leaching tests for regulatory purposes, arises from the high degree of variability apparent in both the extraction and quantitation methodologies employed in the *in vitro* studies. In simple terms, the TNO and LGC studies do little more than demonstrate that phthalates can be extracted from PVC toys into saliva simulants during agitation, and that the precise agitation conditions employed have a substantial influence on the results obtained. They also suggest that a substantial proportion of the intra- and inter-laboratory variability identified may be inherent to the techniques employed, particularly with respect to the HPLC analytical methodology employed by TNO.

In contrast to the conclusions drawn in the respective studies that the methods have now been validated, the results as presented rather demonstrate that such leaching tests do not provide a reliable and responsible basis for the development and application of regulatory measures. Furthermore, although the potential problems of multiple chemical release from PVC toys is acknowledged in the LGC study, it seems unlikely that a methodology based on the determination of leaching rates (and comparison with doses deemed acceptable) could ever be capable of addressing the complexity of chemical leaching from soft PVC toys.

Altogether, neither the *in vivo* nor the *in vitro* methodologies appear to represent sufficient basis for suitably protective regimes, particularly given that the subject of regulation in this case is the avoidable exposure of young children to chemicals which, when handled in the laboratory, are treated as hazardous chemicals. In each case, the studies rely on estimates of tolerable daily intake for phthalates and, moreover, on the proportion of overall phthalate exposure which may be attributable directly to toys. While information on which the TDI is based is limited, information on which to base judgements regarding the significance of other sources of phthalate exposure is virtually non-existent.

Greenpeace International retains its position, therefore, that the only effective and acceptable protective regime would be to take steps to avoid exposure of young children to phthalates and other chemicals contained in, and leaching from, soft PVC toys rather than to attempt to regulate such exposure within levels deemed to be tolerable. The high degree of variability evident from the *in vivo* and *in vitro* validation studies serve to reaffirm this position. Avoidance of exposure can be achieved simply and most effectively by ensuring that only suitable alternative materials to soft PVC, which do not require leachable chemical additives, are used in the manufacture of teethers and other toys designed for use by young children.

[1]Greenpeace Research Laboratories, Department of Biological Sciences, University of Exeter, Exeter EX4 4PS, UK. [2]Greenpeace International European Unit, 37 rue de la Tourelle, B-1040 Bruxelles, Belgium.

1. TNO Nutrition and Food Research Institute, Netherlands, May 1999 – *in vitro* validation

Synopsis

This study (Rijk & Ehlert 1999) attempted to validate the "Head Over Heels" *in vitro* extraction method developed by TNO through a programme of inter-laboratory calibration and comparison with the *in vivo* extraction rates obtained in the unpublished RIVM study. Five laboratories were involved in the validation of extraction and quantitative analytical techniques for diisononylphthalate (DINP) only. The study concludes that the *in vitro* extraction method was "suitable for the intended purpose and can be applied with good repeatability and acceptable reproducibility".

Critique

1. The study employed HPLC for the quantitation of DINP, rather than higher resolution gas chromatography techniques. Using HPLC, DINP resolves as a single peak, making quantitation simpler but losing vital information on possible variations in isomeric complexity between products. It is also noted that DINP and DIDP co-elute using HPLC, a characteristic which would seem to compromise the utility of the technique as these two phthalates may be found together in some products.

2. Values for repeatability of recovery of DINP from a standard solution are presented for only 4 of the 5 participating laboratories. Moreover, repeatability of the HPLC method was >5% (which may be considered a limit for acceptable laboratory practice) in at least one trial from each of two of these laboratories (Nos 3 & 4). Reproducibility between laboratories was relatively poor (23.2%) for a standard analytical technique applied to standard solutions.

3. Intra-laboratory repeatability of the agitation extraction technique was poorer still (2.5-26.7%). Again, data are reported for only 4 of 5 laboratories. Reproducibility was similarly limited (27.5%). It is, therefore, difficult to see how the technique could be considered to be a validated basis for "the intended purpose".

4. The report states that, in relation to reproducibility, the major error can be attributed to the HPLC method. However, within laboratories, the greater part of the error was introduced by the agitation method. The reason for this discrepancy is unclear.

5. The summary statistics presented for the individual toys in table 13, and cited in the text of the report and its summary, indicate a high degree of variability around mean leaching rates (e.g. inter-laboratory confidence levels of 33-73%). Such variation would appear to suggest that the methodology is unsuitable as the basis for regulation.

6. Moreover, these summary statistics disguise an even greater range of variation apparent from individual laboratories. This is important as routine application of techniques for regulatory purposes is likely to be conducted by individual laboratories. Even excluding data for sample 9 (which appeared to be particularly variable), the degree of variation is large (see table 1).

Sample	Mean (ug/min)	relative confidence limit (%)	range of means from individual laboratories	Lowest mean – 95% c.i. to highest mean + 95% c.i.
2	2.4	39.6	1.0 - 3.7	0.94 - 4.08
4	1.5	34.2	1.1 - 2.1	0.30 - 2.35
6	3.7	49.1	1.7 - 5.4	1.18 - 6.26
8	2.1	36.3	1.2 - 2.9	0.97 - 3.38
10	2.7	49.6	1.5 - 4.0	1.02 - 4.24

Table 1: comparison of summary means with means from individual laboratories. Final column gives the possible range of values at 95% confidence, i.e. the lowest reported mean for an individual laboratory adjusted down by its 95% confidence interval and the highest reported mean adjusted up by its confidence interval.

7. Although the method appears to identify some toys which have relatively high or low leaching rates (compared to each other), confidence intervals are substantial. Taking all laboratories combined, and again excluding sample 9, the confidence ranges for the lowest (sample 11) and highest (sample 6) mean leaching rates reported overlap, i.e.

lowest mean 0.9 ug/min (sample 11); upper confidence estimate (mean + 95% c.i.) 1.55 ug/min
highest mean 3.7 ug/min (sample 6); lower confidence estimate (mean – 95% c.i.) 1.18 ug/min

8. Figure 3 demonstrates the presence of an analyte additional to DINP in sample 2, present at significant levels relative to DINP. No attempt was made to identify or quantify this compound, and no mention is made of any similar compounds which may have eluted from other extracts. In our experience, extracts of PVC toys yield complex chromatograms, often revealing the presence of more than one phthalate and a range other additives and/or contaminants. Although the toxicological significance of these additional compounds is not known, simply ignoring their presence does not appear to be a responsible basis for regulation of exposure to chemicals leaching from PVC toys.

Notwithstanding the technical limitations identified above, consideration of the validation of the study is based on comparison of mean *in vitro* leaching rates achieved for two samples, *nos* 1 (standard disk) and 2 (disk cut from teether) with those obtained for the same articles in the unpublished *in vivo* study conducted by RIVM. This comparison raises further concerns:-

9. despite being conducted some time ago, the RIVM *in vivo* study has still yet to be published and subject to scrutiny by the wider scientific community. It is difficult to see how any technique could be considered to have been validated when this validation is based in part on comparison with unpublished data. Further concerns surrounding the RIVM *in vivo* study are detailed below.

10. the opinion of the CSTEE of November 1998 stated clearly that the maximum release rate observed in the *in vivo* studies, and not the mean rates, should be used as a benchmark for comparison:-

"The interlaboratory comparison should be performed using a target value for DINP release of 9 ug/10cm^2/min, which is the maximum release rate observed in the volunteer studies."

Despite this, the TNO study claims validation based on comparison with mean release rates, albeit adjusted upwards according to their respective confidence intervals.

11. Moreover, it is important to note that the *in vivo* human volunteer study conducted by the U.S. Consumer Product Safety Commission (CPSC 1998) yielded mean and maximum leaching rates for PVC toys which were significantly higher than those recorded in the RIVM study. For example, the highest leaching rate determined in the US study, averaged over 1 hour, was 9.65 ug/10cm^2/min. The same individual yielded a leaching rate over a single 15 minute period of 13.37 ug/min. In this case, even the target value of 9 ug/min, adopted by the CSTEE as a "worst case" leaching rate, may need to be revised upwards

12. Notwithstanding points 10 and 11 above, TNO did not even use the highest mean value recorded in the RIVM study. In the RIVM *in vivo* study, three specimens were included. Of these, specimen 2 was considered to be "the most realistic" (RIVM 1998) and was consequently used for the updating of the risk assessment. Specimen 2 also yielded the highest *in vivo* leaching rates (mean 2.4 ug/min) of the three specimens included. Despite this, it would appear that specimen 2 was not included in the subsequent TNO validation, which selected specimen 1 (now sample 1) and specimen 3 (now sample 2) only. These two specimens demonstrated substantially lower mean *in vivo* leaching rates (1.38 and 1.63 ug/min respectively) than specimen 2 (even though specimen 3 was a disk cut from the same toy as the section used as specimen 2), and yet the TNO report still considers the method to have been validated against the RIVM study. The reason for the exclusion of specimen 2 from the TNO study must be clearly justified.

13. Even despite the selectivity of *in vivo* rates for comparison outlined above, it should also be noted that *in vitro* leaching rates for specimens 1 and 3 (samples 1 and 2 in the TNO report) equated or exceeded the *in vivo* rates in only 4 out of 5 participating laboratories. The significance of this 20% quality control failure is largely ignored by TNO and the method is considered to have been validated on the basis of 80% compliance. This "one in five" failure rate introduces a further source of substantial error which may fatally compromise the use of the *in vitro* technique as a basis for regulation.

Rijk, R., & Ehlert, K. (1999) Validation of the method "Determination of Diisononylphthalate in saliva simulant", TNO Report V99.598, 27th May 1999

CPSC (1998) The risk of chronic toxicity associated with exposure to diisononyl phthalate (DINP) in children's products. Directorate for Epidemiology and Health Sciences, U.S. Consumer Product Safety Commission, December 1998

2. Laboratory of the Government Chemist, UK, June 1999 – *in vitro* validation

Synopsis

This study (Axford *et al.* 1999) also attempted to validate two *in vitro* agitation-based extraction methods, using one standard PVC disc and four PVC toys, and involving six participating laboratories (including LGC). Unlike the TNO study, consideration of validation of the LGC techniques was not based on comparison with *in vivo* leaching rate studies but rather on comparison of maximum leaching rates achieved with the CSTEE target value of 9ug/min. The study focused principally on DINP, DIDP and DEHP extraction, although the potential for the presence and leachability of other organic compounds was mentioned.

Critique

1. the majority of the report focuses on the repeatability and reproducibility of the GC-FID and GC-MS analytical techniques and of the recovery of phthalates from standard solutions. Repeatabilities and reproducibilities are, as may be expected, fairly good in these cases. However, statistics are less favourable for migration trials for the reference PVC disks and the toys (see below). Nevertheless, it is these latter trials which are vital for the determination of the applicability of the agitation extraction method for regulatory purposes.

2. As for the TNO study, repeatabilities and reproducibilities for both analytical and extraction techniques were relatively high. For example, for the GC-MS method, mean intra-laboratory repeatability was 9.5%. Moreover, repeatability and reproducibility for the extraction process at 65'C were as high as 22.7% and 31.1% respectively. Again it is difficult to see how the conclusion that such values are "acceptable" within the context of a test for regulatory purposes, was reached

3. Only 5 of the laboratories participated in the calibration of the agitation test with the reference discs at both 37 and 65'C. At 65'C, 3 of the labs ran 6 replicates, 1 ran 5 replicates and 1 only 3 replicates. The difference between replicate numbers is not explained and the consequences for statistical analysis not considered further.

4. The mean leaching rates obtained at 65'C for the individual labs were as follows (abstracted from table 5b):-

Lab Code	3	4	5	6	8
Replicates	6	3	6	5	6
Mean (ug/10cm^2/min)	9.07	13.02	9.06	6.22	8.46
Std Dev. (ug/10cm^2/min)	1.18	2.87	2.18	2.21	1.89
CV%	12.99	22.04	24.08	35.61	22.32

Mean leaching rates determined by individual laboratories therefore varied from 6.22±2.21 ug/10cm^2/min to 13.02±2.87 ug/10cm^2/min, for the same standardised PVC material. Although such variation may appear less significant when the data from all laboratories are taken together, such discrepancies are very highly significant in relation to the comparability of standardised tests which might ultimately be employed by individual laboratories as part of any regulatory process.

5. with respect to the 4 commercial items tested under the same conditions (65'C), the results also demonstrated a high degree of intra- and inter-laboratory variation (abstracted from table 6, all rates again as ug/10cm^2/min):-

Sample	Participating laboratory				
	1	3	5	6	8
Fruit teether	5.30	2.99	5.23	5.57	5.98
Inflatable bath animal	not detected	4.79	interference	2.42	4.48
Swimming armband	1.98	4.48	interference	1.59	4.18
Highchair material	2.86	5.3	interference	3.76	4.36

The coefficient of variation varied from 23.3% (for the teether) to 48.5% (for the armband). Note also that these coefficients do not take account of the non-detects or inability to quantify because of interference. Failure to take into account the non-detect in calculating the mean, standard deviation and coefficient of variation is a substantial statistical error which gives the impression of a lower degree of variation than was actually the case.

6. Despite the high variability which is clear from the report, the discussion concludes the following:-

"The repeatablity and reproducibility of the results obtained demonstrates that both the laboratory-based linear horizontal shaking waterbath agitation methods with GC-MS analytical detection are suitable for the intended purpose of routine enforcement for determining migration of phthalate esters from PVC toys and childcare articles."

Nevertheless, the report lists a number of fundamental limitations to the validation process:-

- *"This is despite the results representing a relatively limited and shortened exercise in terms of normal practice for method validation by collaborative trial. However, this was necessary to provide interested parties with valid methodology as soon as practically possible."*

- *"In addition, the number of participating laboratories means that full use of statistical analysis, in terms of accounting for outliers and stragglers, could not be justified."*

These statements imply that the level of control which would normally be exercised in the case of method validation was relaxed in this case, and that the need for rapid development of a method took precedent over its proper validation. Given that the intention is for the use of this method not merely for regulatory purposes, but ostensibly for the protection of children's health, this is a serious breach of laboratory practice.

- *"In the majority of cases, therefore, the results represent laboratories' first attempts at the methods. This undoubtedly explains the observed differences between individual laboratories and it is clearly foreseeable that the repeatability and reproducibility of the methods will be improved on their repeated use."*

No grounds are given for this assumption, which must be considered to be speculative only.

- *"Any variability in the individual toys has not been taken into account and the mechanism of release of phthalate plasticisers from various material types has been assumed to be similar."*

As for the TNO study, on the basis of reliance on these assumptions and limitations, and the substantial variation in leaching rates recorded both within and between laboratories, it is difficult to see how the conclusion that the method had been properly validated could be supported. Moreover, the extension of the argument that the method could, on the basis of these results, be used for the development of regulations is similarly unwise.

Axford, I.P., Earls, A.O., Scott, R.P., & Braybrook, J.H. (1999). Interlaboratory validation of laboratory-based agitiation methods for the determination of phthalate plasticiser migration from PVC toys and childcare articles. Laboratory of the Government Chemist Technical Report LGC/1999/DTI/004: 24 pp.

3. RIVM/Dutch Consensus Group, Netherlands, September 1998 – *in vivo* study

Synopsis

This study (RIVM 1998), conducted by the Dutch Institute RIVM on behalf of the Dutch Consensus Group, included an experiment to determine the leaching rate of phthalates from standardised PVC discs and two different sections from a PVC hand-shaped toy during sucking and chewing by human volunteers. Also included was a child observation study, conducted in order to revise the estimated mean and maximum contact times with PVC toys for different age groups. Some findings from the study were published as part of the final report of the Dutch Consensus Report in September 1998 (RVIM Report 613320 002), but complete data sets still await publication.

Critique

The discussion below is restricted to the elements of the study relating to the *in vivo* study and are necessarily limited to consideration of the summary data published in the Consensus Group Report.

1. the RIVM study focused entirely on DINP, based on the observation that this is the predominant phthalate in PVC toys. This immediately constrains any conclusions which may be drawn regarding the release of phthalates and other organic compounds from the toys.

2. interpretation of the *in vivo* study was founded on a number of fundamental assumptions which either have not or cannot be validated. For example:-

- *"Differences in composition of saliva of children and adults might be more significant* [than intra-individual variation in composition of saliva], *but quantitative information was not available and such differences were therefore ignored"*.

- *"Because these opposite influences* [i.e. absorption through the membrane of the oral cavity and failure to swallow all of the saliva produced] *could not be quantified, the consensus group decided to ignore them"*.

While it is clear that quantifying these influences would be extremely difficult, if not impossible, simply ignoring their potential significance and giving them no further consideration represents a rather irresponsible approach.

3. the *in vivo* study employed 20 human volunteers. Of these, all 20 tested the standardised PVC disc. The other two specimens, different sections cut from identical PVC toys, were tested by 10 volunteers each. Unfortunately, only the summary statistics for DINP release are presented in the RIVM report, including mean release rate and maximum and minimum values (abstracted from Table 1 of that report):-

Specimen	description	volunteers	mean release rate (ug/min)	95% c.i.	Range (min – max)
1	Standard PVC disc	20	1.38	0.38	0.3 - 8.3
2	Hand toy (finger)	10	2.44	nd	0.9 - 8.9
3	Hand toy (disc)	10	1.63	0.45	0.9 - 5.7

It is not possible to evaluate the data further on the basis of these summary statistics only. Nevertheless, it is clear from the ranges presented that the variation between volunteers was substantial and was not properly reflected by the use of the mean ± 95% confidence interval. For example, for specimen 3, the mean plus 95% c.i. (2.08 ug/min) substantially underestimates the leaching rates recorded for at least one of the 10 volunteers which tested this specimen (maximum 5.7 ug/min). For the other specimens, the degree of underestimation appears to be greater still.

In short, the data set for the *in vivo* study was too small and heavily skewed to be adequately represented by the simple summary statistics of the mean and confidence interval. It is these statistics which, however, are the basis for consideration of the validation of the TNO *in vitro* study (see above). Indeed, none of the laboratories participating in the *in vitro* study generated leaching values which even approach the maximum leaching values recorded for the limited number of volunteers in the RIVM study. It is vital that the full data set from the RIVM human volunteer study be published in order for these limitations to be evaluated further.

4. As in any study which attempts to calculate exposures from measured leaching rates, the doses calculated depend heavily on default values and assumptions for a number of different factors, particularly body weight, exposure time and the percentage of overall phthalate exposure which may be attributable to exposure from toys. For example, for the 6-12 month old category, the RIVM study assumed a body weight of 9.25±1 kg and a proportion of overall exposure attributable to toys of 66.6% (i.e. in allowing 0.1 mg/kg/day of the TDI of 0.15 mg/kg/day to be accounted for by exposure to toys). In contrast, the study conducted by Austrian researchers (Steiner *et al.* 1998) assumed a longer exposure time and, more importantly, only 20% of total exposure attributable to toys.

These differences have substantial influence on the interpretation of the results obtained from the volunteer study. According to the RIVM leaching data, exposure to an unacceptable level (>0.1 mg/kg/day) would be expected only in <1% of cases. However, the same leaching rates would lead to exceedence of the 0.03 mg/kg/day dose deemed acceptable by the Austrian study in >5% of cases. This is not withstanding, of course, the potential significance of even a <1% exceedence when considering the total population of 6-12 month olds over the whole of Europe.

In this regard, the near absence of data relating to exposures of phthalates from sources other than PVC toys remains a substantial and unavoidable limitation to such calculations.

5. The data on which revision of the exposure time from a default of 3 hours to a lower value was based were also very limited, involving a small number of participating children in each age group. While this has no impact on the validation of the TNO *in vitro* study, the validity of this exposure study must ultimately also undergo peer review by the wider scientific community.

In short, the *in vivo* study which has subsequently been used as the basis for validation of the TNO *in vitro* method is actually a very limited study, involving a small number of volunteers, necessitating numerous unvalidated assumptions and reporting substantial intra-individual

variation (with ranges for leaching rate from the standard reference PVC disc covering greater than one order of magnitude). High variation in individually recorded *in vivo* leaching rates was also noted in the Austrian study (Steiner *et al.* 1998).

RIVM (1998) Phthalate release from soft PVC baby toys: Report from the Dutch Consensus Group. Konemann, W.H., [Ed.] National Institute of Public Health and the Environment, Netherlands, RIVM report 613320 002, September 1998: 29pp.

Steiner, I, Kubesch, K., & Fiala, F., (1998) Preliminary Summary of the study "Migration of DEHP and DINP from PVC articles", 3rd September 1998