

Review Article

Information derived from sensitization test methods: test sensitivity, false positives and false negatives

DAVID A. BASKETTER¹ AND IAN KIMBER²

¹Safety and Environmental Assurance Centre, Unilever, Bedford, and ²Syngenta Central Toxicology Laboratory, Cheshire, UK

Predictive toxicology tests for the prospective identification of skin-sensitizing chemicals are well known and have been used for many years. However, of these, only the local lymph node assay (LLNA) has actually undergone formal independent assessment to determine the accuracy of the predictions, particularly with respect to the likelihood of false positives and false negatives. Often, efforts to increase the sensitivity of a test (reducing false negatives) tend to increase the number of false positives. In this short review, these issues are discussed in particular relation to the 3 predictive tests available in regulatory toxicology, the guinea-pig maximization test, the occluded patch test of Buehler and the LLNA. A key perspective is that no predictive test is without limitations; having a good appreciation of these limitations is necessary for making the best use of the information derived from these methods.

Key words: Buehler test; false negative; false positive; guinea-pig maximization test; local lymph node assay; skin sensitization. © Blackwell Munksgaard 2007.

Accepted for publication 20 September 2006

Any toxicological test must fulfil some basic requirements to be of practical value in the protection of human health. Tests must be relevant to the end-point of concern (in this case, skin sensitization) and be reliable, that is, reproducibly able to give the correct prediction when the test is repeated, either at the same laboratory or in different locations. The process by which this requirement is confirmed and demonstrated for new/*in vitro* tests is called 'validation', and this activity has been very clearly prescribed (1, 2).

Tests for the prospective identification of chemicals that possess the ability to cause significant skin sensitization have been available for more than half a century [reviewed in Andersen and Maibach, Botham et al., Basketter and Steiling et al. (3–6)]. These methods have normally used the guinea pig as the species of choice, following on from the early studies in the past century (7). Of the many methods described, none of which was formally validated, ultimately only the guinea pig maximization test (GPMT) of Magnusson and Kligman (8) and the occluded patch test of Buehler (9) have continued to be accepted. Relatively recently, an alternative model using the mouse, the local lymph node assay (LLNA), has gained widespread acceptance following formal

validation and acceptance in the USA and in Europe (10, 11).

For all the methods available currently, it is necessary to have an understanding of the data that are generated, how they can be interpreted and the likelihood that they are correct, in other words, to have an appreciation of the risk that the result may be either a false positive or a false negative. In this brief commentary, the questions of relative test sensitivity, false positives and false negatives in the 3 *in vivo* assays for the predictive identification of skin sensitization hazard are addressed.

Test Sensitivity

Following the earliest development of guinea-pig methods, which were of limited sensitivity, first Buehler (9, 12) and then Magnusson and Kligman (8) proposed protocols designed to offer sufficient sensitivity to identify weaker, as well as strong, skin sensitizers. All these assays employed dermal routes of exposure and occlusion during the induction phase, followed by a topical application challenge to show whether any degree of skin sensitization had been induced. The occluded patch test reported by Buehler, however, was originally

not sufficiently well described and proved very susceptible to technical variations which could severely compromise the assay. This led to the impression that the test was not particularly sensitive and, in part, prompted Magnusson and Kligman to report very thoroughly the work they had done to establish a highly sensitive GPMT (8). Properly conducted, the Buehler test proved adequately sensitive and capable of detecting those skin sensitizers which would require classification according to EU regulations (13, 14).

However, in these laudable efforts to enhance test sensitivity, little consideration was given to the matter of specificity, in other words, the risk of increasing the number of false positives at the expense of reducing the numbers of false negatives (15). Furthermore, much of this occurred at a time when there was no requirement for validation, independent or otherwise, of any protocol. This is of considerable importance, however, as enhancement of sensitivity at the expense of selectivity inevitably leads to over-classification of substances and an associated lowering of the value of the classification; a failure to identify potential sensitizers in contrast can lead to a lack of risk assessment and risk management implementation. Currently, there has been no formal review of the guinea pig tests in this respect, although the view persists that the Buehler test offers greater specificity, whereas the GPMT provides greater sensitivity.

In recent years, the LLNA has been widely adopted (16, 17). In this assay, sensitizing activity is measured as a function of the cell proliferation triggered in lymph nodes draining the site of epicutaneous application of the test chemical. This assay was subject to the full rigour of independent validation, with 200+ chemicals being used to show its sensitivity and specificity (10, 16). In this process, the assay was shown to approach 90% specificity and sensitivity, a view confirmed when an expanded data set was examined (18).

False Positives

No predictive toxicology test is without limitations. Incorrectly identifying a chemical with very limited/no sensitizing activity (for humans) as positive is bound to happen. This might occur for a variety of reasons, including inter-species differences or the fact that the end-point of the predictive test is not a direct mechanistic correlate of human biology. For the guinea pig tests of lesser sensitivity, the false positive results were rarely reported as a practical problem. However, for the highly sensitive GPMT, false positives were considered more likely to occur, eventually with

Kligman himself describing the problem and potential solutions (15). One of the first described GPMT false positive results reported was sulphamonic acid (19). This chemical was correctly identified as non-sensitizing in the LLNA, but this assay reported the classic irritant sodium lauryl sulfate as a false positive (20). The characterization of false positivity in each of these cases rested heavily on human experience, demonstrating that despite extensive skin exposure, neither of these chemicals had been found to sensitize humans. Of potential importance in this respect may be the requirement within the GPMT for exposure to test chemical intradermally in combination with a powerful adjuvant.

An additional issue that is worth mentioning here is that the term 'false positive' is often misapplied in respect of materials which, although actually sensitizers, are widely regarded as insufficiently sensitizing to merit concern (e.g. regulatory classification) but which in testing are found to be sufficiently positive to classify. An example here could be isopropyl myristate, regarded as a false positive in the LLNA (21) but nevertheless associated with occasional clear evidence of being a human contact allergen and thus, strictly speaking, not a true non-sensitizer (22). A contrasting, but comparable, issue arises with false negatives (see below).

False Negatives

In this situation, substances that in reality should be recorded as skin sensitizers fail to cause a (sufficient) response in the test system. To a great extent, this was the driving force behind the development of the GPMT. Thus, false negatives may be less common in this method, but they still exist, an obvious example being the paraben family of preservatives, which are well described human sensitizers; although, a key question is whether they are really of sufficient sensitizing power to merit formal classification (23). Paraben is also negative in other predictive assays; it contrasts with isopropyl myristate (see above), which is a similarly weak sensitizer, but which emerges as a clear, if very weak, sensitizer in the LLNA (21). Often, apparent false negatives may be the result of poor quality test conduct, e.g. with methyldibromoglutaronitrile (MDGN) (23). Early published studies in the guinea pig indicated that this substance was not a significant skin sensitizer (24, 25). Subsequently, human clinical data contradicted these observations, suggesting that MDGN was capable of inducing skin sensitization. Overall, the picture has been of a substance which clearly possessed significant skin sensitization potential.

Most notable was the work of Wilkinson et al. who collated data over a 10-year period for many European clinics and demonstrated a sharp rise in the frequency of positive patch test reactions to MDGN (26). Many of the cases of sensitization to MDGN have been demonstrated to be associated with normal cosmetic and occupational use of this preservative (27, 28). Even the short-term exposure associated with repeated use of liquid soaps preserved with MDGN is a significant cause of MDGN allergic dermatitis (29). This preservative was subsequently shown to be a strong sensitizer in the LLNA (21, 30). Such divergent results may arise simply as a consequence of the many technical challenges of conducting the guinea pig test methods, where factors such as the efficacy of fur removal and quality of patch occlusion can have a dramatic impact on test outcome (5). The issue of technical variations impacting the outcome of guinea pig tests was clearly identified many years ago (4, 5) and led to the introduction of the requirement for weak to moderately sensitizing positive controls in an updated OECD guideline for skin sensitization (31) and the newer guideline for the LLNA (32). It is our view that these are essential for the demonstration of proper test performance, but where a laboratory has appropriate experience, they need not be conducted in parallel with each study (33). Published evidence shows that the response to positive controls in the LLNA is highly reproducible (34).

Guidance on Test Interpretation

It is not intended to provide extensive details/case studies here on how to interpret studies. However, there is considerable information in the literature to which the reader is referred. In the specific case of the 2 guinea pig tests, then weak reactions occurring at challenge may well be best assessed by the conduct of a rechallenge (under identical conditions) 2 weeks later on the alternate flank (15, 35). The guiding principle here is that true sensitization is generally reproducible, whereas non-specific reactions such as skin irritancy are not. More general guidance on the identification of false positives and negatives in both guinea pig and mouse assays has been published (20). Specifically for the LLNA, attention has also been drawn to the importance of considering historical vehicle control data (36). Further examples and considerations have most recently been presented in relation to EU classification considerations, remembering there that the aim is to distinguish substances which should be regarded as skin sensitizers from substances which are either non-sensitizers or are not sufficiently sensitizing to

merit formal classification (30). Ultimately, the interpretation of test data will always have to be undertaken on a case-by-case basis.

Conclusions/Future Directions

If properly conducted, the current portfolio of predictive skin sensitization methods, and notably the well-validated LLNA, provide a reliable screen for contact allergens. Informed deployment of positive controls and of information from historical vehicle control data in the LLNA or of rechallenge data in guinea-pig tests are a valuable aid to data interpretation. In addition, the LLNA also provides valuable information on potency, so that risk assessment/management can be properly implemented on a sound evidential basis (21, 37–41). Obtaining equivalent information using a combination of *in silico* and/or *in vitro* procedures presents substantial challenges (42).

References

1. Zeiger E, Stokes W S. Validating new tests for regulatory acceptance. *Regul Toxicol Pharmacol* 1998; 27: 32–37.
2. Worth A P, Balls M. The principles of validation and the ECVAM validation approach. *Altern Lab Anim* 2002; 30 (Suppl 2): 15–21.
3. Andersen K E, Maibach H I. *Contact Allergy Predictive Tests in Guinea Pigs. Current Problems in Dermatology*, Vol. 14, Basel, Karger, 1985.
4. Botham P A, Basketter D A, Maurer Th, Mueller D, Potokar M, Bontinck W J. Skin sensitization – a critical review of predictive test methods in animal and man. *Food Chem Toxicol* 1991; 29: 275–286.
5. Basketter D A. Guinea pig predictive test for contact hypersensitivity. In: *Immunotoxicology and Immunopharmacology*, Dean J, Munson A, Luster M Kimber I (eds), New York: Raven Press, 1994: pp. 693–702.
6. Steiling W, Basketter D A, Berthold K et al. Skin sensitisation testing – new perspectives and recommendations. *Food Chem Toxicol* 2001; 39: 293–301.
7. Landsteiner K, Jacobs J L. Studies on the sensitization of animals with simple chemicals III. *J Exp Med* 1936; 64: 625–639.
8. Magnusson B, Kligman A M. *Allergic Contact Dermatitis in the Guinea Pig. Identification of Contact Allergens*. Springfield, Illinois, Charles C Thomas, 1970.
9. Buehler E V. Delayed contact hypersensitivity in the guinea pig. *Arch Dermatol* 1965; 91: 171–177.
10. NIH. *The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds*. NIH No. 99–4494, 1999.
11. Balls M, Hellsten E. Statement on the validity of the local lymph node assay for skin sensitisation testing. ECVAM Joint Research Centre, European Commission, Ispra. *Altern Lab Anim* 2000; 28: 366–367.
12. Buehler E V. A rationale for the selection of occlusion to induce and elicit delayed contact hypersensitivity in the guinea pig. A prospective test. *Curr Probl Dermatol* 1985; 14: 39–58.
13. Basketter D A, Gerberick G F. Interlaboratory evaluation of the Buehler test. *Contact Dermatitis* 1996; 35: 146–151.

14. EEC Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Off J Eur Comm* 1967; 196: 1–98.
15. Kligman A M, Basketter D A. A critical commentary and updating of the guinea pig maximisation test. *Contact Dermatitis* 1995; 32: 129–134.
16. Gerberick G F, Ryan C A, Kimber I, Dearman R J, Lea L J, Basketter D A. Local lymph node assay validation assessment for regulatory purposes. *Am J Contact Dermat* 2000; 11: 13–18.
17. Cockshott A, Evans P, Gerberick G F, Betts C J, Dearman R J, Kimber I, Basketter D A. Use and abuse of the local lymph node assay: a regulatory perspective. *Hum Exp Toxicol* 2006; 25: 387–394.
18. Basketter D A, Scholes E W, Evans C D, Kimber I. Divergent responses to sulphanylic acid in the guinea pig maximization test and the local lymph node assay. *Contact Dermatitis* 1992; 27: 209–213.
19. Wright Z W, Basketter D A, Dearman R J, Warbrick E V, Ryan C A, Gerberick G F, Kimber I. The local lymph node assay: further comparisons with guinea pig and human sensitization data. *Toxicol Sci* 2001; 60: 171.
20. Basketter D A, Gerberick G F, Kimber I. Strategies for identifying false positive responses in predictive sensitisation tests. *Food Chem Toxicol* 1998; 36: 327–333.
21. Basketter D A, Kimber I. Predictive test for irritants and allergens and their use in quantitative risk assessment. In: *Contact Dermatitis*, 4th edition, Frosch P J, Menné T, Lepoittevin J-P (eds): Heidelberg, Springer Verlag, 2006: pp. 179–188.
22. Uter W, Schnuch A, Geier J, Lessmann H. Isopropyl myristate recommended for aimed rather than routine patch testing. *Contact Dermatitis* 2004; 50: 242–244.
23. Gerberick G F, Ryan C A, Kern P S, Schlatter H, Dearman R J, Kimber I, Patlewicz G, Basketter D A. Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives. *Dermatitis* 2005; 16: 157–202.
24. Bruze M, Gruvberger B, Agrup G. Sensitization studies in the guinea pig with the active ingredients of Euxyl K 400. *Contact Dermatitis* 1988; 18: 37–39.
25. Hausen B M. The sensitizing potency Euxyl K 400 and its components 1,2-dibromo-2,4-dicyanobutane and 2-phenoxy-ethanol. *Contact Dermatitis* 1993; 28: 149–153.
26. Wilkinson J D, Shaw S, Andersen K E et al. Monitoring levels of preservative sensitivity in Europe. A 10 year overview (1991–2000). *Contact Dermatitis*. 2002; 46: 207–210.
27. Bruze M, Goossens A, Gruvberger B. Recommendation to include methyldibromo glutaronitrile in the European standard patch test series. *Contact Dermatitis* 2005; 52: 19–23.
28. Jensen C D, Johansen J D, Menné T, Andersen K E. Methyldibromoglutaronitrile in rinse-off products causes allergic contact dermatitis: an experimental study. *Br J Dermatol* 2004; 150: 90–95.
29. Zachariae C, Rastogi S, Devantier C, Menné T, Johansen J D. Methyldibromo glutaronitrile: clinical experience and exposure based risk assessment. *Contact Dermatitis* 2003; 48: 150–154.
30. Basketter D A, McFadden J, Evans P, Andersen K E, Jowsey I. Identification and classification of skin sensitizers: identifying false positives and false negatives. *Contact Dermatitis* 2006; 55: 268–273.
31. Organisation for Economic Cooperation and Development. *Test Guideline 406: Skin Sensitisation*, Paris, OECD, 1992.
32. Organisation for Economic Cooperation and Development. *Test Guideline 429: The Local Lymph Node Assay*, Paris, OECD, 2002.
33. Basketter D A, Evans P, Fielder R J, Dearman R J, Kimber I. Local lymph node assay – validation, conduct and use in practice: a commentary. *Food Chem Toxicol* 2002; 40: 593–598.
34. Dearman R J, Wright Z M, Basketter D A, Ryan C A, Gerberick G F, Kimber I. The suitability of hexyl cinnamic aldehyde as a calibrant for the using local lymph node assay. *Contact Dermatitis* 2001; 44: 357–361.
35. Frankild S, Basketter D A, Andersen K E. The value and limitations of rechallenge in the guinea pig maximisation test. *Contact Dermatitis* 1996; 35: 135–140.
36. Basketter D A, Wilson K, Gilmour N J, Ullman L, Gerberick G F, Ryan C A, Dearman R J, Kimber I. Utility of historical vehicle control data in the interpretation of the local lymph node assay. *Contact Dermatitis* 2003; 49: 37–41.
37. Basketter D A, Blaikie L, Dearman R J et al. Use of the local lymph node assay for the estimation of relative contact allergenic potency. *Contact Dermatitis* 2000; 42: 344–348.
38. Felter S P, Robinson M K, Basketter D A, Gerberick G F. A review of the scientific basis for default uncertainty factors for use in quantitative risk assessment of the induction of allergic contact dermatitis. *Contact Dermatitis* 2002; 47: 257–266.
39. Felter S P, Ryan C A, Basketter D A, Gerberick G F. Application of the risk assessment paradigm to the induction of allergic contact dermatitis. *Regul Toxicol Pharmacol* 2003; 37: 1–10.
40. Basketter D A, Clapp C, Jefferies D, Safford R J, Ryan C A, Gerberick G F, Dearman R J, Kimber I. Predictive identification of human skin sensitisation thresholds. *Contact Dermatitis* 2005; 53: 260–267.
41. Basketter D A, Andersen K E, Lidén C, van Loveren H, Boman A, Kimber I, Alanko K, Berggren E. Evaluation of the skin sensitising potency of chemicals using existing methods and considerations of relevance for elicitation. *Contact Dermatitis* 2005; 52: 39–43.
42. Jowsey I, Basketter D A, Westmoreland C, Kimber I. A future approach to measuring relative skin sensitising potency. *J Appl Toxicol* 2006; 26: 341–350.

Address:
Dr David A. Basketter
Safety and Environmental Assurance Centre
Unilever
Colworth House
Sharnbrook
Bedfordshire MK44 1LQ
UK
Tel: 01234264776
e-mail: david.basketter@unilever.com