Comparison of 17 Methods of Predicting the Carcinogenicity of 30 Chemicals

Bristol et al. (4) recently introduced and summarized 17 sets of predictions of the carcinogenicity of 30 chemicals currently undergoing long-term rodent cancer bioassay in the National Toxicology Program (NTP). These predictions will eventually be compared with the bioassay outcomes. To that end, a summary table was provided by Bristol et al. (4). In the legend to that summary table, Bristol et al. (4) noted that reference should be made to the original papers for a completely reliable representation of the individual predictions made. The summary table is rather complex in construction and, in some aspects, it is unclear. For example, the code “NP” can mean two distinct things, and the separate terms “undecided,” “equivocal,” “no prediction made,” and “abstained” are not always distinct in meaning. Given that this table will eventually be used to check the validity of the predictive methods used, it should be clear and reliable. Toward this end, the table has been simplified below (Table 1). Designations of genotoxic and nongenotoxic have been removed and three terms—“not possible,” “uncertain,” and “equivocal”—have been combined into a question mark (?; not equivalent to a prediction of equivocal carcinogenicity). It will be necessary to review and correct the entries in the table so that an accurate compila- tion of the predictions is available for future reference. For example, it is difficult to reconcile the entries under COMPACT with the data presented by Lewis et al. in their table 3 (2). Further, the reason for having identical entries under SHE and under Kerckaet et al. [the source of the single set of SHE data (5)] is unclear. It is important to have these and other uncertainties clarified to prevent confusion later, a task that can only be done by those who hold the predictions originally submitted (i.e., Bristol et al.).

Several general points are evident from Table 1, each of which may require refinement when a revised summary table is produced, but each of which should essentially survive that process. First, all of the chemicals are credited with at least one positive prediction—no uniformly positive or negative predictions have been made. The least number of positive predictions was for compound 27, where only the CASE system predicted activity. At the other extreme, 14 positive predictions were made for compound 3. Second, there is a wide range of predictions made by these several methods for the 30 chemicals. Thus, according to Table 1, the COMPACT method predicts activity for only 5 of the 25 chemicals considered (20%), and the DEREK system predicts activity for only 7 of the 30 chemicals considered (23%). At the other extreme, Bootman predicts that

![Table 1](image-url)
of 30 chemicals will be carcinogenic (66%). Third, some methods have a high proportion of unresolved predictions, for example, 13 of 17 for the CSWG method and 11 of 17 for the FALS method. Fourth, the carcinogenicity of inorganic chemicals seems to be particularly difficult to predict (e.g., 11/17 unresolved predictions for molybdenum trioxide and 7/17 for both gallium arsenide and vanadium pentoxide). Finally, Lewis et al. (2) list a full set of Ames test results for these 30 chemicals, 8 of which are reported to be mutagenic. It would be useful if those data could be endorsed and made part of the overall study, the Ames test being the most generally used carcinogenicity prediction technique.

Broad analyses, such as that undertaken above, can be done in the absence of the final carcinogenicity data, and they are to be encouraged subsequent to a reformulation of the summary table. With the wide range of performance characteristics evident, it is inevitable that some of the predictive methods will be found to be of no general value when the final cancer data are available. Whether any of the assays that perform well can be generally adopted for routine use will then become the leading question. The hope must be that such decisions will be faced and that some methods will disappear while others will be generally endorsed and developed. What is certain is that a major shake-out of predictive methodologies lies ahead of us, as illustrated by the need to understand why there is only 60% agreement between the predictions made by Tennant and Spalding (4) and those made by Ashby (5)—each of whom used the same predictive method (6).

John Ashby
Zenea Central Toxicology Laboratory
Alderley Park, Cheshire, United Kingdom

REFERENCES

1. Bristol DW, Wachman JT, Greenwell A. The NIEHS Predictive-Toxicology Evaluation Project. Environ Health Perspect 104(suppl 5):1001−1010 (1996).
2. Lewis DFV, Ioannides C, Parke DV. COMPACT and molecular structure in toxicity assessment: a prospective evaluation of 30 chemicals currently being tested for rodent carcinogenesis by the NCI/NTP. Environ Health Perspect 104(suppl 5):1011−1016 (1996).
3. Kerckaert GA, Brauninger R, LeBoeuf RA, Isoft RJ. Use of the Syrian hamster embryo cell transformation assay for carcinogenicity prediction of chemicals currently being conducted on 25 organic chemicals. Environ Health Perspect 104(suppl 5):1075−1084 (1996).
4. Tennant RW, Spalding J. Predictions for the outcome of rodent carcinogenicity bioassays: identification of trans-species carcinogens and noncarcinogens. Environ Health Perspect 104(suppl 5):1095−1100 (1996).
5. Ashby J. Prediction of rodent carcinogenicity for 30 chemicals. Environ Health Perspect 104(suppl 5):1101−1104 (1996).
6. Ashby J, Tennant RW. Prediction of rodent carcinogenicity for 44 chemicals: results. Mutagenesis 9:7−15 (1994).

Response

The primary goal of the NIEHS Predictive-Toxicology Evaluation (PTE) project is to stimulate active involvement of researchers from a wide range of disciplines, in both the development and evaluation of predictive toxicology methods. This goal dictates that its project managers perform their roles as objectively as possible. Thus, to announce a new PTE experiment, we use only the information needed to unambiguously identify those chemical bioassays that make up the test set. To support the efforts of participants, we strive to distribute or otherwise provide access to all chemical and toxicological information that we are aware of and which may be of some value to them. However, in the course of managing these aspects of the PTE project, we try meticulously to avoid interjecting inferences about information or features that we or others might judge to hold the most promise for developing or improving predictive-toxicology methods. To do otherwise would impede the generation of creative ideas urgently needed in this challenging research area.

When we wrote the introduction (1) for the 13 PTE2 manuscripts that were published in Environmental Health Perspectives, Supplement 5, our purpose in compiling the wealth of information embodied by the 17 sets of predictions was to provide an overview of state-of-the-art ideas, methodologies, and techniques employed by those who participated in the PTE2 experiment. To be consistent with the primary goal for the PTE project, we compiled tables so as to minimize the influence they might have on the evaluation and interpretation of method performance. We hope that leaving the way open stimulates others to engage these difficult issues. Judging from John Ashby’s response, our approach is producing valuable results.

We encourage comments and suggestions like those presented by John Ashby, and we hope to receive them from many more people interested in predictive toxicology research. The spectrum of suggestions received will be incorporated into a publication that evaluates the state of the science and art of predictive toxicology. The simplified table offered by John Ashby has the virtue of clarity. It is useful for quick comparison of the overall performance of the different predictive methods and it enables one to keep score easily. However, the gain in clarity is at the expense of information that is important to evaluating other aspects of method performance, such as, which methods can predict an equivocal response or which methods apply to a broad range of chemical-substance types, including mixtures, etc. Perhaps a series of tables with decreasing information content is needed to guide evaluation of all results from the PTE2 experiment.

As the 30 test-set bioassays of PTE2 near completion, we plan to conduct a workshop to analyze performance of the various predictive methods, identify their strengths, and evaluate all aspects of the PTE experiment. The workshop will focus on answering consequential questions such as the following: What have we learned? Are there gaps in data or other impediments to the development, confirmation, and utilization of models that can be eliminated or ameliorated? What are the most promising directions for further research and development? Certainly, all concerns raised about performance evaluation will be considered at that time. Tables will be prepared ahead of the workshop to help guide its evaluations. However, to resolve the ambiguities and uncertainties inherent to the compilation process, we will obtain input and seek concurrence from the authors who generated the prediction sets. The need for this is illustrated by the closing comment in John Ashby’s letter. It asks why agreement is not better between two prediction sets that were generated using the same method. Ashby’s observation is an interesting early result from the PTE2 experiment. It suggests that the role of intuition in the application of implicit rules by human experts is more significant than previously estimated. This conjecture is supported by practical experience, which shows that the bottleneck to building expert computer systems is the excessive time required to extract and refine implicit rules from knowledge-domain experts before they can be converted into machine-friendly, explicit form (2). Yet, who other than Ashby, Tennant, and Spalding is in a better position to evaluate the differences, answer the questions, and perhaps improve the method?

Douglas W. Bristol
Joseph T. Wachman
Arnold Greenwell
National Institute of Environmental Health Sciences
Research Triangle Park, North Carolina

REFERENCES

1. Bristol DW, Wachman JT, Greenwell A. The NIEHS Predictive-Toxicology Evaluation Project. Environ Health Perspect 104(suppl 5):1001−1010 (1996).
2. Quinlin JR. Simplifying decision trees. Int J Man−Mach Stud 27:221−234 (1987).