



COMPACT Predictions: Is There a Catch?

Lewis et al. reported a retrospective evaluation of COMPACT predictions of rodent carcinogenicity for 44 chemicals evaluated in long-term bioassays by the National Toxicology Program (*EHP* 103:178–184). They concluded that COMPACT performed quite well. I was surprised to read this, because I had seen the published COMPACT carcinogenicity predictions regarding these chemicals (*1*) and knew that the method had not performed particularly well. Thus, I was interested in learning how the method's predictive performance had been enhanced. After examining the paper by Lewis et al. (*1*) in more detail, it quickly became clear that the authors had employed several questionable data manipulations in their retrospective analysis to improve the performance of their method.

Publishing predictions of carcinogenicity before the study outcomes are known firmly establishes the predictions and permits an easy assessment of their accuracy. Prospective predictions are important because a predictive methodology that truly works should be able to predict the carcinogenic potential of untested chemicals. Unfortunately, COMPACT had little success in this regard. Based on Lewis et al.'s predictions (*1*), COMPACT was able to correctly predict the carcinogenicity outcome for only 56% (20/36) of the NTP chemicals, a success rate not significantly different from flipping a coin. The 16 chemicals for which these COMPACT predictions were inaccurate are given in Table 1.

Given these results, it would have been appropriate for the authors to attempt to understand the reasons behind COMPACT's failures and to make the necessary modifications so that the method might be more successful when applied prospectively to a new set of chemicals. Instead, the authors reevaluated the predictions made for the 44 NTP chemicals and manipulated the data in various ways to show that COMPACT (when used in combination with Hazardexpert) really had only 5 discordant carcinogenicity predictions, not 16. The more important data manipulations are discussed below.

Retrospective changes in the COMPACT predictions. For two chemicals, Lewis et al. changed their previously published carcinogenicity predictions from positive to negative. For HC Yellow 4, the authors stated in footnote *b* to Table 1 (p.

Table 1. Chemicals for which COMPACT incorrectly predicted carcinogenicity

NTP noncarcinogens predicted to be carcinogens	NTP carcinogens predicted to be noncarcinogens
Promethazine	<i>o</i> -Benzyl- <i>p</i> -chlorophenol
Resorcinol	Methylphenidate hydrochloride
<i>p</i> -Nitrophenol	Diphenylhydantoin
Tricresyl phosphate	Tris(2-chloroethyl)phosphate
Chloramine	2,3-Dibromo-1-propanol
4,4'-Diamino 2,2'-stilbenedisulfonic acid	1,2,3-Trichloropropane
CI Pigment Red 23	
4-Hydroxyacetanilide (acetaminophen)	
HC Yellow 4	
<i>p</i> -Nitroaniline	

Table 2. COMPACT and Hazardexpert predictions for six NTP equivocal carcinogens

Chemical	COMPACT prediction	Hazardexpert prediction	Carcinogenicity outcome ^a
γ -Butyrolactone	-	-	-
Chloramine	+	-	-
CI Pigment Red 23	+	+	+
4-Hydroxyacetanilide	+	-	+
HC Yellow 4	-	+	-
<i>p</i> -Nitroaniline	+	+	+

^aThese carcinogenicity outcomes reflect the views of Lewis et al., not the conclusions of the NTP. Lewis et al. concluded that the carcinogenicity outcomes for all six chemicals were predicted correctly by COMPACT/Hazardexpert.

179) that while their original prediction was positive, "calculation based on new structure gives negative." For resorcinol, the authors apparently justify the changed prediction by asserting that "the original graphical analysis was clearly negative." However, the paper they cite to justify the negative graphical analysis is their previous paper (*1*), in which they clearly report their carcinogenicity prediction for resorcinol to be positive, not negative. Both changes result in correct predictions, reducing the number of discordant predictions from 16 to 14.

Reinterpretation of NTP's equivocal carcinogenicity results. Equivocal responses often occur in rodent carcinogenicity studies. Lewis et al. state that for assessing concordance "when this single [the carcinogenicity] response is EE (equivocal evidence), the overall response is . . . taken as '-' in the final assessment" (pp. 178–179). This is a reasonable approach, but the authors did not follow this rule when assessing the predictive performance of COMPACT.

There were nine NTP chemicals for which only equivocal evidence of carcinogenicity was observed. For three equivocal carcinogens predicted by COMPACT to be positive (CI pigment red 23, 4-hydroxyacetanilide, and *p*-nitroaniline), the authors reevaluated the results and concluded that these studies were "weak positives/equivocal positives based on patholo-

gy reports" (Table 2, p. 180), and thus their predictions that the chemicals would be carcinogens were correct after all. This reduced the number of discordant predictions from 14 to 11.

The source of the "pathology reports" is not given, but clearly Lewis et al.'s interpretation of these studies does not reflect the views of the NTP, which concluded that these three bioassays showed equivocal, not weakly positive, carcinogenic effects (as did the other six chemicals showing equivocal responses). There are several reasonable options for dealing with equivocal carcinogenicity outcomes. One is to regard them all as positive or all as negative (the latter being the rule the authors claim to have followed, as noted previously). Alternatively, chemicals with equivocal or uncertain findings could be excluded from consideration altogether when evaluating predictive methodologies. Less defensible is the strategy used by the authors, who attempted to distinguish between "equivocal positives" (which they considered positive) and "equivocal negatives" (which they considered negative).

COMPACT predicted carcinogenicity outcomes for six of the nine NTP equivocal carcinogens, and these predictions are summarized in Table 2, together with the Lewis et al. interpretation of the carcinogenicity outcomes. The authors concluded that all six chemicals are predicted correctly by COM-

PACT/Hazardexpert. However, by the rule they claim to have used (equivocal carcinogens are regarded as noncarcinogens), only γ -butyrolactone is predicted correctly.

Inclusion of additional related variables in the predictive method. After the carcinogenicity outcomes were known, Lewis et al. found that the predictive performance of COMPACT could be enhanced if they included an additional COMPACT prediction (C2E) and also a predictor variable ("Hazardexpert") that incorporates information about metabolism. While there is nothing inherently wrong with including additional variables in a predictive methodology, this exercise should ideally have been carried out prospectively, not retrospectively. It is much easier to find predictive variables that work once the study outcomes to be predicted are known. The important (and yet to be answered) question is, how will the authors' newly derived, multivariate predictive methodology fare for prospective predictions? Hopefully, it will be better than COMPACT's limited predictive success (56%) for the 44 NTP chemicals.

The combination of COMPACT and Hazardexpert eliminated the apparent discordance for three chemicals: tris(2-chloroethyl)phosphate, 2,3-dibromo-1-propanol, and 1,2,3-trichloropropane, while introducing discordance for another chemical previously predicted correctly (methyl bromide). However, the authors misclassify two other chemicals: chloramine and HC Yellow 4, both of which are reported as successful predictions, but in fact were not predicted correctly (see Table 2). Including additional predictor variables (and not correcting for the misclassification of chloramine and HC Yellow 4) reduced the number of discordant predictions from 11 to 8.

Inclusion of additional, apparently unrelated, variables in the predictive method. The eight chemicals that Lewis et al. conclude are not correctly predicted by COMPACT/Hazardexpert are designated in their Table 4 (p. 182). The authors then carry out further analyses to reduce the number of discordant predictions from eight to five. Frankly, it is unclear exactly how the authors achieve this reduction. It appears that the basis for eliminating the final three chemicals from "discordancy" was an appeal to "structural alert, chronic toxicity studies, and the Ames test," which correctly predicted the carcinogenicity of *o*-benzyl-*p*-chlorophenol, methylphenidate hydrochloride, and diphenylhydantoin, three chemicals "missed" by COMPACT/Hazardexpert. One other chemical (mercuric chloride) not even evaluated by COMPACT/Hazardexpert, but correctly identified by "the metal ion redox poten-

tials for inorganic compounds," was also apparently added in as a correct prediction. The authors should justify how these additional predictions, based on apparently unrelated variables, can be meaningfully interpreted as improving the performance of COMPACT/Hazardexpert. In any case, the authors include these successful predictions in their calculations and conclude that the concordance for COMPACT/Hazardexpert when predicting rodent carcinogenicity is 86% (32/37). I leave it to the reader's judgment to determine how much confidence to place in this figure.

I have no objection to the development of techniques designed to predict rodent (or more importantly, human) carcinogenicity, and I suspect that it is possible to develop methods that will be successful in this regard. However, I strongly urge caution in placing too much confidence in COMPACT or in any other predictive method that has little success when applied prospectively and seems to work only when applied retrospectively to the original data set, using extensive (and scientifically questionable) data manipulations and reanalysis.

Joseph K. Haseman

National Institute of Environmental Health Sciences
Research Triangle Park, North Carolina

REFERENCE

1. Lewis DFV, Ioannides C, Parke DV. A prospective toxicity evaluation (COMPACT) on 40 chemicals currently being tested by the National Toxicology Program. *Mutagenesis* 5:433-435 (1990).

Response

In response to Joe Haseman's letter, we would like to point out that although our article is retrospective with regard to the rodent carcinogenicity study of the 40 chemicals, the COMPACT data were available at the time of the release of the carcinogenicity assays. The Hazardexpert evaluations for the 40 chemicals were carried out after the NIEHS conference, but the Hazardexpert system (available commercially from Compudrug Ltd) is not part of COMPACT. The following account, hopefully, provides some clarification of the points raised in Dr. Haseman's letter.

Most other systems publish their predictions or analyses without providing any mathematical derivation which can be reproduced by others. In contrast, we show how our predictions/analyses are generated from numerical values (COMPACT parameters) for molecular and electronic features of each chemical. Our attempts to provide a numerical description of the

COMPACT plot of molecular planarity/potential chemical reactivity, have not been entirely successful, due to the fact that the training set of chemicals shows a curved line discriminating P4501 specificity from other P450 isozymes, whereas the COMPACT ratio (either $\text{area/depth}^2/\Delta E$ or $\text{area/depth}^2/\Delta E - 8$) gives a straight line relationship. This results in some chemicals (e.g., resorcinol) having a COMPACT ratio and a COMPACT graphical plot which give conflicting results, but the graph is the *original* paradigm. We have recently derived an expression which is more complex (1), based on analysis of the COMPACT curve, and this gives more precise results in terms of correlation with the graph, although the actual graphical representation is preferred.

Resorcinol was predicted to be positive in COMPACT using the COMPACT ratio, but the graph of area/depth^2 versus ΔE as presented at the 1993 NTP conference (2) clearly shows that this compound should be negative as it is outside the curve. This is the only example in all of the 40 chemicals of a discrepancy between the approximation of the COMPACT ratio and the accurate description of the graph. As the *EHP* paper is retrospective, we feel justified in making this point, even though the graphical description was available in the conference documentation. HC Yellow 4 was changed from positive in COMPACT to negative, due to the fact that the original structure sent to us by NTP was erroneous and was subsequently changed by NTP after our original predictions had been published. When we ran the new (correct) structure through our system, it proved negative, and we feel justified in making this clear in our retrospective study published in *EHP*. However, we provided revised data (including the aforementioned cases) and distributed this at the NTP conference, which, moreover, included our results for the P4502E descriptor, now provided in the February 1995 issue of *EHP* (103:178-184).

The Hazardexpert analyses were generated retrospectively as we had only recently purchased the software. As can be seen from our *EHP* paper, the Hazardexpert results (which utilize the EPA database) give quite good concordances with positive carcinogens, and they are better than the Ames test for negatives and also overall.

Regarding the relatively poor performance of the original computer-based predictions compared with that of Ashby, Tennant, and others, it should be emphasized that the latter employed a combination of mutagenicity, subchronic toxicity, and structural alert tests, which are, therefore, three evaluations combined into one prediction—so it is perhaps not surprising