

The Humane Society Institute for Science and Policy

## WBI Studies Repository

2-2006

### Animal Carcinogenicity Studies: 1. Poor Human Predictivity

Andrew Knight

*Animal Consultants International*

Jarrold Bailey

*University of Newcastle-upon-Tyne*

Jonathan Balcombe

*Physicians Committee for Responsible Medicine*

Follow this and additional works at: [https://www.wellbeingintludiesrepository.org/acwp\\_arte](https://www.wellbeingintludiesrepository.org/acwp_arte)



Part of the [Animal Experimentation and Research Commons](#), [Animal Studies Commons](#), and the [Other Medical Sciences Commons](#)

---

#### Recommended Citation

Knight, A., Bailey, J., & Balcombe, J. (2006). Animal carcinogenicity studies: 1. Poor human predictivity. *ATLA-NOTTINGHAM*, 34(1), 19.

This material is brought to you for free and open access by WellBeing International. It has been accepted for inclusion by an authorized administrator of the WBI Studies Repository. For more information, please contact [wbisr-info@wellbeingintl.org](mailto:wbisr-info@wellbeingintl.org).



# Animal Carcinogenicity Studies: 1. Poor Human Predictivity

Andrew Knight,<sup>1</sup> Jarrod Bailey<sup>2</sup> and Jonathan Balcombe<sup>3</sup>

<sup>1</sup>Animal Consultants International, London, UK; <sup>2</sup>School of Population and Health Sciences, Faculty of Medical Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne, UK; <sup>3</sup>Physicians Committee for Responsible Medicine, Washington DC, USA

**Summary** — The regulation of human exposure to potentially carcinogenic chemicals constitutes society's most important use of animal carcinogenicity data. Environmental contaminants of greatest concern within the USA are listed in the Environmental Protection Agency's (EPA's) Integrated Risk Information System (IRIS) chemicals database. However, of the 160 IRIS chemicals lacking even limited human exposure data but possessing animal data that had received a human carcinogenicity assessment by 1 January 2004, we found that in most cases (58.1%; 93/160), the EPA considered animal carcinogenicity data inadequate to support a classification of probable human carcinogen or non-carcinogen. For the 128 chemicals with human or animal data also assessed by the World Health Organisation's International Agency for Research on Cancer (IARC), human carcinogenicity classifications were compatible with EPA classifications only for those 17 having at least limited human data ( $p = 0.5896$ ). For those 111 primarily reliant on animal data, the EPA was much more likely than the IARC to assign carcinogenicity classifications indicative of greater human risk ( $p < 0.0001$ ). The IARC is a leading international authority on carcinogenicity assessments, and its significantly different human carcinogenicity classifications of identical chemicals indicate that: 1) in the absence of significant human data, the EPA is over-reliant on animal carcinogenicity data; 2) as a result, the EPA tends to over-predict carcinogenic risk; and 3) the true predictivity for human carcinogenicity of animal data is even poorer than is indicated by EPA figures alone. The EPA policy of erroneously assuming that tumours in animals are indicative of human carcinogenicity is implicated as a primary cause of these errors.

**Key words:** animal experiment, animal test, bioassay, cancer prevention, carcinogenicity, chemical classification, chemical safety, risk assessment.

**Address for correspondence:** A. Knight, Animal Consultants International, 91 Vanbrugh Court, Wincott Street, London SE11 4NR, UK.  
E-mail: [info@animalconsultants.org](mailto:info@animalconsultants.org)

## Introduction

Since the first chemical bioassay occurred in 1915, when Yamagiwa and Ichikawa showed that coal tar applied to rabbit ears caused skin carcinomas (1), several thousand have been conducted, with the objective of determining the risks posed to humans by the great majority of chemicals for which adequate human exposure data are lacking. However, despite the heavy reliance of governmental regulatory agencies on animal carcinogenicity testing, this remains a controversial area of animal research.

Proponents of the bioassay claim that all known human carcinogens that have been studied in sufficient animal species have produced positive results in one or more species (2–4). Critics respond that, if sufficient animal testing is conducted, carcinogenesis will eventually occur in *some* species, regardless of the cancer risk from a particular chemical. A study published in the journal, *Mutagenesis*, found that of 20 known human *non*-carcinogens, 19 were known animal carcinogens (5). Other investigators have also reported the poor human specificity (the ability to identify human *non*-carcinogens) of car-

cinogenicity bioassays, and have noted the considerable biological and mathematical complexities of attempting to accurately extrapolate animal carcinogenicity data to humans (6–8).

Other key disadvantages of animal carcinogenicity studies are their protracted time-frames, and their substantial demands on human, animal and financial resources. Monro and MacDonald (9) estimated that rodent bioassays take at least three years to plan, execute and interpret. Millions of skilled personnel hours have been consumed in these studies to date.

Ashby (10) estimated that the bioassay evaluations of 400 chemicals by the US National Toxicology Program (NTP) from the 1970s to the 1990s cost hundreds of millions of dollars. As of 2005, the bioassay results of 6153 experiments on 1485 chemicals were included in the comprehensive Berkeley-based carcinogenic potency database (CPDB; 11). Greek and Greek (12) estimated that the cost of carcinogenicity bioassays exceeds 250 million dollars annually.

Similarly, millions of animal lives have been consumed by animal carcinogenicity studies. Monro

and MacDonald (9) estimated that a single carcinogenicity bioassay may use over 1200 animals. Furthermore, data from the USA (13) and Canada (14) indicate that testing procedures such as carcinogenicity studies, account for most of the animals that are reported as experiencing the highest levels of pain and distress in laboratory studies. That pain and distress is not short-term. As exemplified by the US National Cancer Institute/National Toxicology Program (NCI/NTP) protocol, dosing in the standard rodent bioassay begins at six to eight weeks of age and continues for 90 to 110 weeks, a period similar to the natural rodent lifespan, after which any remaining survivors are killed and autopsied (15).

In 2004, Pound *et al.* (16) pointed out in the *British Medical Journal* that justifications for the use of animal tests in safeguarding human health have sometimes relied on anecdotal evidence or unsupported claims. Given increasing concern about the ethical issues posed by animal testing, and increasing competition for scarce research resources, critical reviews of the value of animal tests in safeguarding human health are clearly warranted.

Consequently, we have systematically reviewed the human utility of animal carcinogenicity data for regulatory purposes. The control of human exposure to various potential carcinogens constitutes the most important use of animal carcinogenicity data. The US federal agency most responsible for regulating exposure to potentially dangerous environmental contaminants is the Environmental Protection Agency (EPA; 17), and the chemicals of greatest concern within the USA (18) are listed in the EPA's Integrated Risk Information System (IRIS) chemicals database, along with their toxicity data and human carcinogenicity assessments (19).

To assess the reliability of EPA carcinogenicity assessments, we compared them with those of the World Health Organisation's International Agency for Research on Cancer (IARC), as published in its *IARC Monographs Programme on the Evaluation of Carcinogenic Risks to Humans*. Compiled by international working groups of scientific experts, the *IARC Monographs* provide critical reviews and evaluations of the evidence relating to the possible carcinogenicity of a wide variety of agents, mixtures and exposures. They are recognised as authoritative sources of information, and assist governmental agencies in making risk assessments and in formulating decisions concerning any necessary preventive measures. A 1998 users' survey indicated that the *IARC Monographs* are consulted by various agencies in 57 countries — 4000 copies of each volume are usually printed, for distribution to governments, regulatory bodies and interested scientists (20).

## Methods

We examined the 543 chemicals catalogued in the EPA's IRIS chemicals database (as of 1 January 2004; 21) to determine the proportion for which the EPA was able to derive classifications of probable human carcinogen or probable human non-carcinogen, based primarily on animal carcinogenicity data. The relatively few classifications of definite human carcinogen relied primarily on available human exposure data. The remaining classifications of unclassifiable or possible human carcinogen were not considered substantially useful for risk assessment or regulatory purposes, and are excluded from the NTP's authoritative annual *Report on Carcinogens* (22).

Of the 235 chemicals assigned human carcinogenicity classifications by the EPA, we determined the proportion in each of the following categories, along with the reason for the classification:

- A: Human Carcinogen (convincing evidence of human carcinogenicity).
- B1: Probable Human Carcinogen (limited evidence of human carcinogenicity).
- B2: Probable Human Carcinogen (sufficient evidence of animal carcinogenicity).
- C: Possible Human Carcinogen (animal data inadequate for stronger classification).
- D: Unclassifiable (animal data inadequate for stronger classification).
- D: Unclassifiable (no animal or human data).
- E: Evidence of Non-Carcinogenicity for Humans (sufficient evidence of non-carcinogenicity, at least in animals).

The 160 chemicals lacking even limited human carcinogenicity data but possessing animal carcinogenicity data, were then examined, to determine the proportion for which the EPA considered the animal data strong enough to assign the classifications of probable human carcinogen (B2) or probable human non-carcinogen (E). A 95% confidence interval for this proportion was derived via the modified Wald method, which is described in *The American Statistician* as providing more accurate results than the so-called "exact" method commonly used (23).

To compare EPA carcinogenicity classifications with those assigned by the IARC, we examined the 885 agents (chemicals, groups of chemicals, complex mixtures, occupational exposures, cultural habits, biological or physical agents) assigned human carcinogenicity classifications in the first 82 volumes of the *IARC Monographs* series published by 1 January 2004. We determined the proportion in each of the following IARC categories:

- 1: Human Carcinogen.

- 2A: Probable Human Carcinogen.
- 2B: Possible Human Carcinogen.
- 3: Human Carcinogenicity Unclassifiable.
- 4: Probable Human Non-Carcinogen.

Like the EPA, the IARC classified definite human carcinogens on the basis of convincing human data, and probable human non-carcinogens on at least sufficient animal data. Unlike the EPA, however, the IARC did not subdivide probable human carcinogens into those based on limited evidence of human carcinogenicity and those based solely on animal carcinogenicity data. This prevented calculation of the proportion of agents classified by the IARC as probable human carcinogens primarily on the basis of their animal carcinogenicity data, other than by examination of a large number of IARC agents individually, an approach which we did not undertake. This is one of the reasons we chose instead to use EPA classifications to derive our initial assessments of the human utility of animal carcinogenicity data.

Of the 177 chemicals that had received a human carcinogenicity classification from the EPA based on human or animal data, 128 were assigned human carcinogenicity classifications by both the EPA and the IARC. These 128 were divided into those considered by the EPA to possess at least limited human data (17 chemicals) and those primarily reliant on animal data (111 chemicals) for their human carcinogenicity classifications.

The consistency of classifications between the EPA and IARC was examined for these two groups, by comparing the carcinogenicity classification proportions within each group via chi-squared tests, and by comparing the individual classifications of the 111 chemicals primarily reliant on animal carcinogenicity data for their human carcinogenicity classifications.

Chi-squared tests provide a statistical calculation of the probability that two data sets, such as the EPA and IARC human carcinogenicity classifications, are samples from the same underlying data population, and that any observed differences are simply due to random sampling variation. Large chi-squared values ( $\chi^2$ ) reflect increased probabilities that observed differences are due to real differences in underlying data populations.

Chi-squared and two-tailed  $p$  values were derived from the online statistical calculators available at <http://www.graphpad.com/quickcalcs/index.cfm>.

## Results

### EPA and IARC human carcinogenicity classifications

Of the 543 chemicals catalogued in the EPA’s IRIS chemicals database, 235 had been assigned human carcinogenicity classifications. Of these, 17 were classified as definite (A) or probable (B1) human carcinogens on the basis of their human carcinogenicity data. Of the remaining 218 chemicals lacking even limited human data, 160 were deemed to possess animal carcinogenicity data (B2, C, subset of D, and E; Table 1).

Table 2 lists the IARC human carcinogenicity classifications of the 885 agents described in the first 82 volumes of the *IARC Monographs* series. As stated, the IARC also classified definite human carcinogens (category 1) on the basis of convincing human data, and probable human non-carcinogens (category 4) required at least sufficient animal data. Unlike the EPA classifications of B1 and B2, however, the IARC did not subdivide probable human carcinogens (2A) into those based on limited evidence of human carcinogenicity, and those based solely on animal carcinogenicity data.

**Table 1: EPA human carcinogenicity classifications of IRIS chemicals**

EPA human carcinogenicity classification	Number of chemicals	% of total
A: Human Carcinogen (convincing human data)	11	4.7
B1: Probable Human Carcinogen (limited human data)	6	2.6
B2: Probable Human Carcinogen (sufficient animal or human data)	64	27.2
C: Possible Human Carcinogen (animal data inadequate for stronger classification)	40	17.0
D: Unclassifiable (animal data inadequate for stronger classification)	53	22.6
D: Unclassifiable (no animal or human data)	58	24.7
E: Probable Human Non-Carcinogen (sufficient animal data)	3	1.3
Total	235	

*160 chemicals lacking in human data had received a human carcinogenicity assessment primarily on the basis of their animal data. Data source: EPA Integrated Risk Information System database, 1 January 2004.*

**Table 2: IARC human carcinogenicity classifications of agents published in the IARC Monographs**

IARC human carcinogenicity classification	Number of chemicals	% of total
1: Definite Human Carcinogen	88	9.9
2A: Probable Human Carcinogen	64	7.2
2B: Possible Human Carcinogen	237	26.8
3: Unclassifiable	495	55.9
4: Probable Human Non-Carcinogen	1	0.1
Total	885	

Data source: IARC Monographs Programme on the Evaluation of Carcinogenic Risks to Humans, Volumes 1–82, 1 January 2004.

### The human utility of animal carcinogenicity data based on EPA figures

Of the 160 EPA chemicals lacking even limited human data (A or B1) but possessing animal data (B2, C, subset of D, and E), 64 were considered probable human carcinogens (B2), and three were considered probably not carcinogenic to humans (E). For those considered probably not carcinogenic to humans, in some cases, data arising from human exposure or mechanistic knowledge contributed to the assessment. The remaining 93 chemicals were considered unclassifiable as to their human carcinogenicity (D; 53) or to be possible human carcinogens (C; 40), based on animal data considered inadequate to support a stronger classification (Table 1).

Overall, of those 160 chemicals lacking even limited human data but possessing animal data, the EPA considered the animal data inadequate to support the substantially useful classifications of probable human carcinogen or probable human non-carcinogen in the majority of cases (93/160; 58.1%, 95% CI: 50.4–65.5).

### Comparison of EPA and IARC human carcinogenicity classifications

Of those 177 chemicals that had received a human carcinogenicity classification from the EPA based on human or animal data (A, B1, B2, C, D with animal data, or E), 128 were also assessed by the IARC. Of these, 17 were considered by the EPA to possess at least limited human data (A or B1), and the remaining 111 EPA carcinogenicity classifications were primarily reliant on animal data.

For those 17 chemicals considered by the EPA to possess at least limited human data, overall EPA classifications were not found to differ significantly from those predicted by IARC classifications ( $\chi^2 = 0.291$ ,  $df = 1$ ,  $p = 0.5896$ ; Table 3). N.B. Chi-squared analysis does not allow comparison when one category lacks

any data. Hence acrylonitrile, assessed as the only possible human carcinogen by the IARC, but as a probable human carcinogen (B1) by the EPA, was excluded, yielding a more conservative result.

However, for those 111 chemicals considered by the EPA to lack even limited human data, but to possess animal data, EPA and IARC classifications were very significantly different overall ( $\chi^2 = 215.548$ ,  $df = 2$ ,  $p < 0.0001$ ; Figure 1). To permit chi-squared analysis, methacrylate, assessed as unclassifiable by the IARC, but as the only probable human non-carcinogen by the EPA, was excluded, yielding a more conservative result.

The data reveal that the EPA was much more likely than the IARC to assign carcinogenicity classifications indicative of greater human hazard. The numbers of chemicals classified by the EPA as probable human carcinogens (60 chemicals) compared to all other categories (51 chemicals) were very significantly different from those predicted by the IARC, for which the equivalent numbers of chemicals were 12 and 99 ( $\chi^2 = 215.273$ ,  $df = 1$ ,  $p < 0.0001$ ). Similar disparities were found for possible human carcinogens ( $\chi^2 = 19.771$ ,  $df = 1$ ,  $p < 0.0001$ ) and unclassifiable chemicals ( $\chi^2 = 24.378$ ,  $df = 1$ ,  $p < 0.0001$ ).

Comparison of the individual classifications of these 111 chemicals revealed that 67 (60.4%) were assigned an EPA carcinogenicity classification indicative of greater human hazard, 38 (34.2%) were assigned an equivalent classification, and 6 (5.4%) were assigned a classification indicative of lower human hazard than the corresponding IARC classification of the same chemical.

## Discussion

### Differing EPA and IARC human carcinogenicity classifications

Based on EPA figures alone, the predictivity of animal carcinogenicity data for human hazard, and

**Table 3: IARC classifications of EPA chemicals possessing significant human data (EPA categories A or B1)**

Human carcinogenicity classification	EPA	IARC
Human Carcinogen (A)	11	12
Probable Human Carcinogen (B1)	6	4
Possible Human Carcinogen	0	1
Total	17	17

Data sources: The EPA Integrated Risk Information System database, 1 January 2004, and the IARC Monographs Programme on the Evaluation of Carcinogenic Risks to Humans, Volumes 1–82, 1 January 2004.

hence its utility in deriving substantially useful human carcinogenicity classifications, is clearly questionable. Of those 160 IRIS chemicals lacking even limited human data but possessing animal data, the EPA considered the animal data inadequate to support the substantially useful classifications of probable human carcinogen or non-carcinogen in the majority (93) of cases. Classifications of definite human carcinogen relied on the existence of convincing human data. Classifications of unclassifiable or possible human carcinogen were not considered substantially useful for risk assessment or regulatory purposes, and are excluded from the NTP’s annual *Report on Carcinogens* (22).

However, IARC assessments of the same chemicals reveal that the human utility of animal carcinogenicity data is probably even lower than indicated by EPA figures. EPA and IARC carcinogenicity classifications were similar only for those chemicals having human data. For those lacking human data, the EPA was much more likely than the IARC to assign carcinogenicity classifications indicative of greater human hazard. Of chemicals lacking human data assessed by both agencies, the EPA classified 61 chemicals as probable human carcinogens or non-carcinogens, primarily on the basis of their animal data. In contrast, the IARC classified only 12 chemicals similarly, assessing the remainder as unclassifiable or as possible human carcinogens.

Given that the IARC is recognised as one of the most authoritative sources of information on potential human carcinogens (20, 24), it is implausible that IARC assessments would generally be inaccurate or based on incomplete data. Consequently, the significant differences in human carcinogenicity classifications of identical chemicals between the IARC and the EPA indicate that:

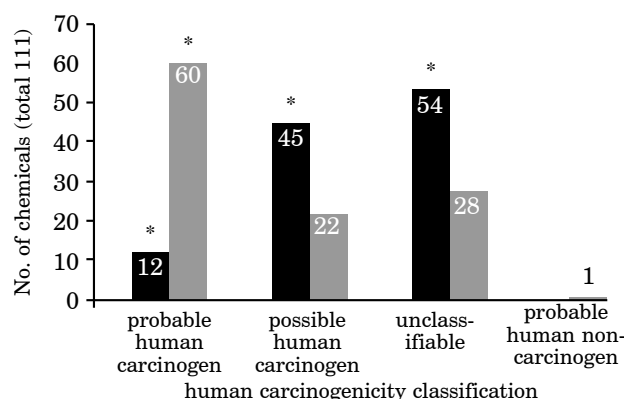
1. in the absence of significant human data, the EPA is over-reliant on animal carcinogenicity data;

2. as a result, the EPA tends to over-predict carcinogenic risk; and
3. the true predictivity for human carcinogenicity of animal data is even poorer than is indicated by EPA figures alone.

**Differences in EPA and IARC human carcinogenicity assessments**

Both the EPA and the IARC include a wide range of data in their carcinogenicity assessments. The

**Figure 1: EPA and IARC human carcinogenicity classifications of chemicals considered by the EPA to lack human data but to possess animal data**



■ = IARC; ■ = EPA; \* = p < 0.0001.  
 Data sources: The EPA Integrated Risk Information System database, 1 January 2004, and the IARC Monographs Programme on the Evaluation of Carcinogenic Risks to Humans, Volumes 1–82, 1 January 2004.

major types of evidence considered include human epidemiological studies and case reports, or, more rarely, randomised trials; and lifetime exposure studies in test animal species. Conventional standardised rodent carcinogenicity studies use at least 50 animals of each sex per dose group in each of three treatment groups and a concurrent control group, and last for 18 to 24 months. Assays in genetically engineered rodents may also be considered, particularly with a view to elucidating the chemical or genetic mechanisms of carcinogenesis. Such rodents may have activated oncogenes introduced (transgenic rodents) or tumour suppressor genes deleted (knockout rodents). Also considered, where available, are supporting data from short-term genotoxicity tests, such as standard bacterial and mammalian *in vitro* and *in vivo* tests, and data describing preneoplastic lesions, tumour pathology, toxicological effects other than cancer, metabolic and toxicokinetic properties, physicochemical parameters, structure-activity relationships, and data concerning analogous biological agents. On occasion, specific additional studies may be commissioned to fill important knowledge gaps. Both agencies aim to include all pertinent data in deriving a weight-of-evidence assessment of human carcinogenicity.

However, three key differences between the approaches of the EPA and the IARC may explain their very significantly different human carcinogenicity classifications for identical chemicals.

1. Not every EPA assessment is conducted with the same scope or depth. The level of detail of an assessment is a matter of management discretion, which, besides considerations of potential human and environmental risk from a suspected carcinogen, also seeks to balance pragmatic considerations such as the time, personnel and resources required for each particular assessment, with those available at the time, as well as the time and cost of generating any new data required. Consequently, the agency's staff often conduct screening assessments to decide whether to invest resources in collecting data for a full assessment. Such screening assessments may be based almost entirely on structure-activity relationships and default assumptions, and more detailed assessments may not occur (25).

IARC assessments, on the other hand, are invariably conducted in depth. Each *IARC Monograph* evaluation results from detailed consideration and discussion by 15 or more internationally-based scientific experts, and the level of evaluation and deliberation is considered by the IARC to be considerably in excess of the usual peer review process used by scientific journals (26).

2. While both agencies prefer data from peer-reviewed sources, the IARC appears to be more

critical about the standard of the data it accepts for use in its carcinogenicity assessments. Only occasionally are data accepted from sources other than the peer-reviewed scientific literature, such as government agency reports that have undergone peer review (26–27). This decreases the likelihood that data from animal studies of poor quality (for example, those with inadequate durations, animal numbers or survival rates) will be included in IARC assessments (28).

3. As leaders of the US federal agency most responsible for protecting Americans from environmental contaminants in the world's most litigious nation, the policy-makers of the EPA understandably err on the side of caution. The impacts on EPA policy are almost inevitable, as illustrated by the EPA *Guidelines for Carcinogen Risk Assessment*, which state that: "The primary goal of EPA actions is protection of human health; accordingly, as an Agency policy, risk assessment procedures, including default options that are used in the absence of scientific data to the contrary, should be health protective. Use of health protective risk assessment procedures as described in these cancer guidelines means that estimates, while uncertain, are more likely to overstate than understate hazard and/or risk." (25).

Such policies have affected EPA carcinogenicity assessments for many years. In response to a Congressional directive regarding EPA appropriations for the fiscal year 2000, the EPA undertook an evaluation of data uncertainty and variability within its IRIS assessments. A representative sample of 16 IRIS assessments were subjected to in-depth evaluation by a panel of six independent experts in the field of human health risk assessment. Among other criticisms, they concluded that, despite being advertised as a quantitative, science-based exercise, where uncertainty existed about the data upon which decisions were based, the classifications of some chemicals were, in fact, more reflective of the EPA policy of favouring classifications indicative of greater human risk (29). As noted by the experts, such a policy is consistent with the EPA's mission to protect public health and the environment by means of conservative limits on human exposure to carcinogens, where doubt remains about true carcinogenic risk.

With respect to the use of animal test data, the EPA *Guidelines* state that: "In the absence of sufficiently scientifically justifiable mode of action information, EPA generally takes public health protective, default positions regarding the interpretation of toxicologic and epidemiologic data: animal tumor findings are judged to be relevant to humans..." (25), and, "tumors observed in animals

are generally assumed to indicate that an agent may produce tumors in humans.” (25).

The EPA is strongly defensive of its position: “The default option is that positive effects in animal cancer studies indicate that the agent under study can have carcinogenic potential in humans... This option is a public health-protective policy, and it is both appropriate and necessary, given that we do not test for carcinogenicity in humans...”. Despite strong indications from numerous investigations (4, 6–8, 30–36), the EPA seems reluctant to acknowledge the extent of error this may incur, and states that: “The extent to which animal studies may yield false positive indications for humans is a matter of scientific debate.” (25).

However, EPA carcinogenicity assessments are not necessarily inferior to those of other US regulatory agencies. In their survey of 350 representative chemicals, Viscusi and Hakes (37) found that the human carcinogenicity assessments of other US regulatory authorities, particularly the Food and Drug Administration and the Occupational Safety and Health Administration, are based even less on an accurate assessment of carcinogenicity data than are those of the EPA.

### Results from IARC *Monographs* surveys

The poor human predictivity of animal carcinogenicity studies was also demonstrated in 1993 by Tomatis and Wilbourn (24), who surveyed the 780 chemical agents or exposure circumstances evaluated and listed within Volumes 1–55 of the *IARC Monograph* series (38). Of these, 502 (64.4%) were classified as having definite or limited evidence of animal carcinogenicity, and 104 (13.3%) as definite or probable human carcinogens. Virtually all of the latter group would, of course, have been members of the former; however, around 398 animal carcinogens were considered not to be definite or probable human carcinogens.

The positive predictivity of a test is the proportion of positive test outcomes that are truly positive for the characteristic being tested for, while the false positive rate refers to the proportion that are not. Hence, based on these IARC figures, the positive predictivity of the animal bioassay for definite or probable human carcinogens was only around 20.7% (104/502), while the false positive rate was a disturbing 79.3% (398/502).

More-recent IARC classifications indicate little improvement in the positive predictivity of the animal bioassay for human carcinogens. By 1 January 2004, a decade later, only 105 additional agents had been added to the 1993 figure, yielding a total of 885 agents or exposure circumstances listed in the *IARC Monographs* (39). Not surprisingly, the proportion of definite or probable human carcinogens resembled the 1993 figure of

13.3%. By 2004, only 9.9% of these 885 were classified as definite human carcinogens, and only 7.2% as probable human carcinogens, yielding a total of 17.1% (Table 2).

### Results from NTP and other surveys

Surveys by other investigators have also demonstrated the poor human predictivity of animal carcinogenicity data. After examining the studies of 471 substances contained within the NTP carcinogenicity database as of 1 July 1998, Haseman (30) concluded that although 250 (53.1%) produced carcinogenic effects in at least one sex–species group, the actual proportion posing a significant carcinogenic risk to humans was probably far lower, for reasons such as interspecies differences in mechanisms of carcinogenicity. Similarly, around half of all chemicals tested on animals and included in the comprehensive CPDB, whether natural or synthetic, give positive results (7).

Rall (4) estimated that only around 10% of chemicals are truly carcinogenic to humans. Ashby and Purchase (31) speculated that all chemicals would eventually display some carcinogenic activity, if tested in sufficient rodent strains. Even common table salt has been classified as a tumour promoter in rats (32).

Fung *et al.* (33) estimated that, if all 75,000 chemicals in use were tested for carcinogenicity via the standard NTP bioassay, significantly less than 50% would prove carcinogenic in animals, and less than 5–10% would warrant further investigation. They suggested that the higher positivity rate recorded was due to chemical selection based on an *a priori* suspicion of carcinogenicity. However, examination of the carcinogenicity literature reveals that chemicals are selected for study for many reasons other than an *a priori* suspicion, including production volumes, occupational and environmental exposure risks, and investigations of carcinogenesis mechanisms (34). Despite this, the positivity rate of the carcinogenicity bioassay in the general literature remains around 50% (7).

### Carcinogenicity bioassays fail human validation

Despite the heavy reliance for the last several decades on rodent carcinogenicity data in the regulation of human exposures, the conventional rodent bioassay has never been formally validated against human data. On the contrary, validation studies have found the rodent bioassay to be lacking in human specificity (i.e. in the ability to identify human *non*-carcinogens), resulting in false positive outcomes, or even human sensitivity (i.e. the ability to detect human carcinogens at all), depending on



the data interpretation method used. Ennever and Lave (35) showed that neither of the two commonly-used interpretations of rodent carcinogenicity data provide conclusions about human carcinogenicity that are supported by existing data. If a risk-avoidance interpretation is used, in which any positive result in male or female mice or rats is considered positive, then nine of the 10 known human carcinogens among the hundreds of chemicals tested by the NTP are positive (36), but so are an implausible 22% of all chemicals tested (33). If a less risk-sensitive interpretation is used, whereby only chemicals positive in both mice and rats are considered positive, then only three of the six known human carcinogens tested in both species are positive (36). The former interpretation could result in the needless denial of potentially useful chemicals to society, while the latter could result in widespread human exposure to undetected human carcinogens.

## Conclusions

By 1998, only about 2000 (2.7%) of the 75,000 industrial chemicals in use and listed in the EPA's Toxic Substances Control Act inventory, had been tested for carcinogenicity (40). The cost of testing these 2.7% of industrial chemicals was millions of animal lives (9, 41), millions of hours of work by skilled personnel (41), and hundreds of millions of dollars (10, 12).

The most important use of the animal data thus derived is the regulation of human exposure to potential carcinogens by governmental agencies such as the EPA. However, we found that the human predictivity of animal carcinogenicity data was inadequate for the EPA to derive substantially useful human carcinogenicity classifications for the majority (58.1%) of chemicals studied.

Even when the EPA was able to derive substantially useful human carcinogenicity classifications, primarily on the basis of animal carcinogenicity data, there was a profound discordance with assessments of identical chemicals by the IARC, a leading international authority on carcinogenicity assessments, seriously undermining the reliability of EPA classifications. For those classifications primarily reliant on animal data, the EPA was much more likely than the IARC ( $p < 0.0001$ ) to assign classifications indicative of greater human risk, indicating that: 1) in the absence of significant human data, the EPA is over-reliant on animal carcinogenicity data for extrapolating to human hazard; 2) as a result, the EPA tends to over-predict carcinogenic risk to humans; and 3) the true predictivity for human carcinogenicity of animal data is even poorer than indicated by EPA figures alone.

EPA human carcinogenicity classifications appear less scientifically-based than those of the

IARC, due to: 1) varying depths of EPA assessments, as a result of resource constraints; 2) less rigorous standards required of data incorporated into EPA assessments; and 3) an EPA public health-protective policy that seeks to err on the side of caution by assuming that tumours in animals are indicative of human carcinogenicity.

The sensitivity of the conventional rodent bioassay in detecting human carcinogens for *some* sex-species groups is not in question. However, its poor human predictivity severely limits its utility for assessing human hazard.

## Acknowledgement

This research was partly funded by the Physicians Committee for Responsible Medicine, Washington DC, USA.

Received 6.8.05; received in final form 18.12.05; accepted for publication 19.12.05.

## References

- Huff, J. (1999). Long-term chemical carcinogenesis bioassays predict human cancer hazards. Issues, controversies, and uncertainties. *Annals of the New York Academy of Sciences* **895**, 56–79.
- Wilbourn, J., Haroun, L., Heseltine, E., Kaldor, J., Partensky, C. & Vainio, H. (1986). Response of experimental animals to human carcinogens: an analysis based upon the IARC Monographs Programme. *Carcinogenesis* **7**, 1853–1863.
- Tomatis, L., Aitio, A., Wilbourn, J. & Shuker, L. (1989). Human carcinogens so far identified. *Japanese Journal of Cancer Research* **80**, 795–807.
- Rall, D.P. (2000). Laboratory animal tests and human cancer. *Drug Metabolism Reviews* **2**, 119–128.
- Ennever, F.K., Noonan, T.J. & Rosenkranz, H.S. (1987). The predictivity of animal bioassays and short-term genotoxicity tests for carcinogenicity and non-carcinogenicity to humans. *Mutagenesis* **2**, 73–78.
- Meijers, J.M., Swaen, G.M. & Bloemen, L.J. (1997). The predictive value of animal data in human cancer risk assessment. *Regulatory Toxicology and Pharmacology* **25**, 94–102.
- Gold, L.S., Slone, T.H. & Ames, B.N. (1998). What do animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database. *Drug Metabolism Reviews* **30**, 359–404.
- Monro, A. (1996). Are lifespan rodent carcinogenicity studies defensible for pharmaceutical agents? *Experimental and Toxicologic Pathology* **48**, 155–166.
- Monro, A.M. & MacDonald, J.S. (1998). Evaluation of the carcinogenic potential of pharmaceuticals. Opportunities arising from the International Conference on Harmonisation. *Drug Safety* **18**, 309–319.
- Ashby, J. (1996). Alternatives to the two-species bioassay for the identification of potential human

- carcinogens. *Human and Experimental Toxicology* **15**, 183–202.
11. Gold, L.S., Manley, N.B., Slone, T.H., Rohrbach, L. & Garfinkel, G.B. (2005). Supplement to the Carcinogenic Potency Database (CPDB): results of animal bioassays published in the general literature through 1997 and by the National Toxicology Program in 1997–1998. *Toxicological Sciences* **85**, 747–808.
  12. Greek, C.R. & Greek, J.S. (2000). *Sacred Cows and Golden Geese: The Human Costs of Experiments on Animals*. 242pp. New York, NY, USA: Continuum International.
  13. Stephens, M.L., Mendoza, P., Hamilton, T. & Weaver, A. (1998). Unrelieved pain and distress in animals: an analysis of USDA data on experimental procedures. *Journal of Applied Animal Welfare Sciences* **1**, 15–26.
  14. Canadian Council on Animal Care (1998). *CCAC Animal Use Survey*. 14pp. Ottawa, Canada: CCAC.
  15. Peto, R., Pike, M.C., Bernstein, L., Gold, L.S. & Ames, B.N. (1984). The TD50: A proposed general convention for the numerical description of the carcinogenic potency of chemicals in chronic-exposure animal experiments. *Environmental Health Perspectives* **58**, 1–8.
  16. Pound, P., Ebrahim, S., Sandercock, P., Bracken, M.B. & Roberts, I. (2004). Where is the evidence that animal research benefits humans? Much animal research into potential treatments for humans is wasted because it is poorly conducted and not evaluated through systematic reviews. *British Medical Journal* **328**, 514–517.
  17. Anon. (2004). US Environmental Protection Agency. Website <http://www.epa.gov> (Accessed 29.1.04).
  18. Anon. (2003). *U.S. EPA's Process for IRIS Assessment Development and Review*. Website <http://www.epa.gov/iris/process.htm> (Accessed 10.12.03).
  19. Anon. (2004). *What is IRIS?* Website <http://www.epa.gov/iris/intro.htm> (Accessed 29.1.04).
  20. Anon. (1999). *Objective and Scope*. IARC 7 Dec. 1999. Website <http://www-cie.iarc.fr/monoeval/objectives.html> (Accessed 12.1.05).
  21. Anon. (2004). *IRIS Database for Risk Assessment*. Website <http://www.epa.gov/iris/index.html> (Accessed 1.1.04).
  22. Anon. (2002). *National Toxicology Program Report on Carcinogens, 10th Edn*. Website <http://ntp.niehs.nih.gov/ntpweb/index.cfm?objectId=72016262-BDB7-CEBA-FA60E922B18C2540> (Accessed 13.1.05).
  23. Agresti, A. & Coull, B.A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician* **52**, 119–126.
  24. Tomatis, L. & Wilbourn, J. (1993). Evaluation of carcinogenic risk to humans: the experience of IARC. In *New Frontiers in Cancer Causation* (ed. O. Iversen), pp. 371–387. Washington DC, USA: Taylor and Francis.
  25. Anon. (2005). *Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001B*. Washington DC, USA: Risk Assessment Forum, U.S. Environmental Protection Agency. Website <http://www.epa.gov/iris/backgr-d.htm> (Accessed 11.12.05).
  26. Anon. (2005). *Internal Report 05/001: Report of the Advisory Group to Recommend Updates to the Pre-amble to the IARC Monographs*. Lyon, France: IARC.
  27. Anon. (1999). *Data for the Monographs*. Website <http://www-cie.iarc.fr/monoeval/data.html> (Accessed 12.12.05).
  28. Anon. (2005). *Studies of Cancer in Experimental Animals*. Website <http://www-cie.iarc.fr/monoeval/studiesanimals.html> (Accessed 12.12.05).
  29. Anon. (2000). *Characterisation of Data Uncertainty and Variability in IRIS Assessments: Pre-pilot vs. Pilot/ Post-pilot*. 46pp. Springfield, VA, USA: Versar, Inc.
  30. Haseman, K. (2000). Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk. *Drug Metabolism Reviews* **32**, 169–186.
  31. Ashby, J. & Purchase, I.F.H. (1993). Will all chemicals be carcinogenic to rodents when adequately evaluated? *Carcinogenesis* **8**, 489–495.
  32. Shirai, T., Fukushima, S., Ohshima, M. & Ito, N. (1984). Effects of butylated hydroxyanisole, butylated hydroxytoluene, and NaCl on gastric carcinogenesis initiated with N-methyl-N-nitro-N-nitrosoguanidine in F344 rats. *Journal of the National Cancer Institute* **72**, 1189–1198.
  33. Fung, V., Barrett, J. & Huff, J. (1995). The carcinogenesis bioassay in perspective: application in identifying human hazards. *Environmental Health Perspectives* **103**, 680–683.
  34. Gold, L.S., Bernstein, L., Magaw, R. & Slone, T.H. (1989). Interspecies extrapolation in carcinogenesis: prediction between rats and mice. *Environmental Health Perspectives* **81**, 211–219.
  35. Ennever, F.K. & Lave, L.B. (2003). Implications of the lack of accuracy of the lifetime rodent bioassay for predicting human carcinogenicity. *Regulatory Toxicology and Pharmacology* **38**, 52–57.
  36. Johnson, F.M. (2001). Response to Tennant *et al.*: attempts to replace the NTP rodent bioassay with transgenic alternatives are unlikely to succeed. *Environmental Molecular Mutagenesis* **37**, 89–92.
  37. Viscusi, W.K. & Hakes, J.K. (1998). Synthetic risks, risk potency, and carcinogen regulation. *Journal of Policy Analysis and Management* **17**, 52–73.
  38. International Agency for Research on Cancer (1972–1992). *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. Volumes 1–55. Lyon, France: IARC.
  39. International Agency for Research on Cancer. (undated). *IARC monographs programme on the evaluation of carcinogenic risks to humans*. Website <http://monographs.iarc.fr> (Accessed 1.1.04).
  40. Epstein, S.S. (1998). *The Politics of Cancer, Revisited*. 770pp. Fremont Center, NY, USA: East Ridge Press.
  41. Gold, L.S., Manley, N.B., Slone, T.H. & Rohrbach, L. (1999). Supplement to the Carcinogenic Potency Database (CPDB): results of animal bioassays published in the general literature in 1993 to 1994 and by the National Toxicology Program in 1995 to 1996. *Environmental Health Perspectives* **107**, Suppl. 4, 527–600.