Published in final edited form as:

Environ Sci Technol. 2012 March 6; 46(5): 2538–2544. doi:10.1021/es204310j.

Ratcheting up cancer potency estimates

Edmund A.C. Crouch¹ and Gilbert S Omenn²

Edmund A.C. Crouch: crouch@CambridgeEnvironmental.com; Gilbert S Omenn: gomenn@umich.edu

¹Senior Scientist, Cambridge Environmental Inc., 58 Charles Street, Cambridge, MA 02141;

Office: 617-225-0810; Fax: 617-225-0813; <http://www.CambridgeEnvironmental.com>

²Professor of Internal Medicine, Human Genetics, and Public Health, and Director of the Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, 48109-2218; 734-663-7583; www.ccmb.med.umich.edu/omenn

Summary

The current paradigm for cancer risk assessment in the United States (U.S.) typically requires selection of representative rodent bioassay dose-response data for extrapolation to a single cancer potency estimate for humans. In the absence of extensive further information, the chosen bioassay result generally is taken to be that which gives the highest extrapolated result from the “most sensitive” species or strain. The estimated human cancer potency is thus derived from an upper-bound value on animal cancer potency that is technically similar to an extreme value statistic. Thus additional information from further bioassays can only lead to equal or larger cancer potency estimates. We here calculate the size of this effect using the collected results of a large number of bioassays. Since many standards are predicated on the value of the cancer potency, this effect is undesirable in producing a strong counter-incentive to performing further bioassays.

Keywords

risk assessment; rodent carcinogenesis bioassays; potency; extrapolation to humans; value of information model

Introduction

As this issue of the journal presents tributes to Lester Lave, we have chosen to focus on his well-published interest in the design of efficient and credible testing of chemicals for risk of carcinogenicity for humans. Lave and Omenn¹ and Lave et al.² proposed and utilized a “value of information model” to guide the design of strategies for in vitro assays and for lifetime exposure rodent carcinogenicity assays, respectively. They modeled a range of ratios for the societal cost of false-negatives (inaccurately declaring a carcinogenic chemical to be non-carcinogenic, thus permitting its use and the exposure of some number of people) versus the societal cost of false-positives (inaccurately declaring a chemical to be carcinogenic, thus likely denying or reducing its use). Lave and Omenn also were interested in combining in vitro and in vivo assays to learn more about mechanisms of action of the chemical and to gain a more comprehensive view of its potential hazards. However, in the practical world, manufacturers were reluctant to conduct additional assays due to their impression that such testing would put the chemicals in “double jeopardy” to be declared carcinogenic, at least

some of which cases would be inaccurate. Here we explore another aspect of this phenomenon.

In the report of the Presidential/Congressional Commission on Risk Assessment and Risk Management,³ the Commission reinforced the reliance on rodent lifetime carcinogenicity testing by advocating mechanistic investigations. They also identified several examples in which evidence is overwhelming that the mechanism at play in rodents (often rats or mice but not both) does not occur in humans. Recognizing and removing those exceptions makes the general reliance on extrapolation from rodent bioassays to human lifetime exposure estimates more credible. The Commission then accepted the highly conservative position that, absent a compelling mechanistic explanation for divergent results across species, any chemical that tested positive in either sex or either rodent species should be considered potentially carcinogenic for humans and should undergo appropriate risk management. Like the U.S. regulatory agencies, they seemed to accept the convention that even one positive lifetime rodent carcinogenesis bioassay would override null results in other similar assays.

However, there are policy and statistical reasons why the extrapolation from a highly selected quantitative bioassay result may exaggerate the risk estimate in humans. Such exaggeration is usually justified by a precautionary public safety argument, without taking account of its societal costs. The methodology generally adopted is to select that rodent experiment providing the highest estimate of cancer potency, and extrapolate from that result to humans. We examine here the effect of such an approach on sequential evaluations of the same chemical as further information becomes available, although we caution that nothing examined here demonstrates that the results necessarily lead to overestimates of human cancer potency.

Previous observations

There is a substantial history of analysis of large databases of rodent carcinogenicity bioassays, examining patterns that may be used to empirically justify extrapolations within and between species. The extrapolations examined have included concordance estimates (e.g., Purchase,⁴ Gold et al.,^{5,6} Freedman et al.,⁷) such as used by Lave et al.,² and estimates of the uncertainties in quantitative extrapolation using various measures of carcinogenicity (Crouch and Wilson,⁸ Crouch,⁹ Gaylor and Chen,¹⁰ Chen and Gaylor,¹¹ Allen et al.^{12,13} Crump et al.¹⁴ EPA,¹⁵ Gaylor et al.,¹⁶ and Crouch¹⁷).

Gold et al.¹⁸ amassed a database of a large number of published long-term bioassays conducted on rats, mice, hamsters, dogs, monkeys, and prosimians, although there are too few bioassays on dogs, monkeys, or prosimians for the analysis we perform. The most recent analysis¹⁹ of this database used a multistage dose-response model, and evaluated a point of departure as the lifetime average dose rate (CD₁₀; for Cancer Dose 10%) causing a 10% increment in lifetime cancer risk, to correspond with current practice used in regulatory evaluations of carcinogenicity.²⁰ The CD₁₀ values used here are Benchmark Dose levels calculated from multiple dose points in the dose-response relationship.²¹ The CD₁₀ was used for ease of computation rather than a lower bound on a Benchmark Dose. Similar results are expected using such other measures of cancer potency because of the relatively small

standard deviation (SD) for individual bioassay results compared with the within and between species variation detailed below. Full details of the database analysis are provided by Crouch;¹⁹ that analysis produced the following results (at least for the subset of bioassays that could be included in the analysis) that we use here to evaluate the effect of sequential evaluation of multiple bioassays:

1. For each chemical and species combination, CD_{10} values (evaluated for each bioassay using the end point giving the lowest value that is statistically significant) from different bioassays (of different strains, in different laboratories, at different times) form a lognormal distribution (the within-species, across bioassay, CD_{10} distribution). See Figure 1 for examples. Bioassays on males and females are treated independently, since they appear to be independent in these analyses.
2. The medians of the within-species, across bioassay, CD_{10} distributions for a particular chemical differ between species. The ratio of these medians for combinations of two species (among rat, mouse, and hamster) form between-species lognormal distributions across chemicals (see Figure 2) with medians (Table 1) that do not correspond to any simple allometric scaling rule such as the $\frac{1}{4}$ power of body weight scaling rule that is currently used as standard practice by the U.S. Environmental Protection Agency (EPA), Food and Drug Administration (FDA), and Consumer Product Safety Commission (CPSC) for interspecies extrapolation (this lack of allometric scaling was previously demonstrated by Crouch,¹⁷ using a different measure of cancer potency).
3. The standard deviations of the lognormal within-species, across bioassay, CD_{10} distributions differ between chemicals for the same species ($p < 10^{-100}$, likelihood ratio test), and differ between species for the same chemical ($p < 10^{-14}$, likelihood ratio test).
4. The distribution across chemicals of the standard deviations of the within-species, across bioassay, CD_{10} distributions are indistinguishable from lognormal ($p=0.78$, 0.53 , and 0.61 , Shapiro-Wilk tests, for rat, mouse, and hamster respectively; the Shapiro-Wilk statistics are used here heuristically, since they are constructed from observations each of which has a different associated uncertainty), but these distributions have different parameters for each species (Table 2).
5. The between-species, across-chemical distributions of the difference in means of the within-species, within-chemical normal distributions of the logarithm of CD_{10} have standard deviations (Table 1) that are statistically indistinguishable for rat-mouse, mouse-hamster, hamster-rat, and animal-human (not shown), but this apparent universality is not used here.

Current selection of bioassays

The current approach to selecting bioassay results used to derive values used for cancer potency for regulatory purposes is described in ref. 20. A range of available approaches is described for use when multiple possible values are available, but the method for selecting between such approaches is left open. In practice, as used for example in the EPA's IRIS

database (<http://www.epa.gov/iris/>), the smallest statistically significant CD₁₀ result from the most sensitive species (i.e. the one that gives the largest cancer potency estimate in humans) is generally chosen, although there are exceptions. In fact the empirical results providing the highest potency are generally used without actually proving that the species or strain is, indeed, the most sensitive on a consistent basis either for the chemical or for the end point involved. The rationale is that humans are highly outbred and highly variable in their susceptibility to any toxic agent; if the policy goal is to protect essentially everyone (e.g., to <1 cancer per 100,000 people if exposed for a lifetime at the level of some standard), then extrapolating from the most susceptible rodent is appropriate, assuming that the chosen rodent response has an actual human analogue.

Consequences

Most cancer potency estimates are initially based on at most a few bioassays, very often only two (e.g. on males and females of one species) or four (e.g. those performed by the National Toxicology Program on male and female rats and mice), and these estimates will be made very soon after the bioassay results become available. Such estimates may be used to derive standards, and are certainly used in practice to set either regulatory levels (e.g. Proposition 65 safe harbor limits in California) or “advisory” levels (e.g. soil cleanup standards).

If subsequent bioassays are performed, they can be used to derive different estimates for cancer potency that are either smaller or larger than those initially derived. In the normal course of affairs (and to ensure “conservatism”), the larger ones will replace the initial estimates, while smaller values will essentially be ignored by regulatory and standard setting agencies, except possibly in specific circumstances where sufficient research is available to demonstrate human-relevant mechanistic reasons for the differences.

Numerical examples

Using the parametric distributions of CD₁₀ with parameters obtained empirically from the bioassays summarized by Gold et al.,¹⁸ we demonstrate the expected effect, and the distributions of effects, of additional bioassays on ratcheting up the estimates of cancer potency (equivalently, ratcheting down the estimated CD₁₀ in humans). For simplicity these examples assume that the only bioassays available are in the rat or mouse, and we randomly choose between these species for the next bioassay. Extrapolation from rat or mouse to human is assumed to be performed by dividing the CD₁₀ obtained in rat or mouse by a factor of $3.908 = (70/0.3)^{1/4}$ or $6.950 = (70/0.03)^{1/4}$ respectively, corresponding to allometric scaling by bodyweight with a scaling power of 1/4, the current standard agreed upon by the EPA, FDA, and CPSC. Starting with either 2 or 4 bioassay results, we evaluate the effect on the estimated CD₁₀ in humans of adding 1, 2, 4 or 8 additional bioassays and using the current approach of choosing the lowest CD₁₀ estimate obtained from these as a point of departure for estimating risks to humans.

We used a Monte Carlo simulation of 1,000,000 hypothetical chemicals, which was sufficient to ensure that our results were numerically stable to approximately 3 decimal places. For each chemical, we randomly selected standard deviations for the within- species,

across bioassay CD_{10} distributions for rats and mice from the distributions in Table 2, and randomly selected the difference in medians of these within-species, across bioassay, CD_{10} distributions from the rat versus mouse distribution in Table 1. Without loss of generality (since we compare results only within a single chemical), we set the median CD_{10} for the mouse within-species, across bioassay distribution to unity, so that the within-species, across bioassay median for rat and mouse for this chemical were now selected. We then generated hypothetical bioassay results by first randomly selecting rat or mouse with 50% probability, then randomly choosing a CD_{10} value from the appropriate within-species, across bioassay CD_{10} distribution with the selected median and standard deviation. The required statistics on the minimum CD_{10} after the requisite number of bioassays were generated from these selected CD_{10} values.

As an alternative, we also evaluate the effect of using the geometric mean estimate of all bioassays performed so far. Given the observed distributions of CD_{10} , this metric would be expected to be considerably more stable than the extreme value statistic currently in use.

Results

The probabilities for reductions in estimates of human CD_{10} after one, two, four, or eight additional bioassays of the same chemical are shown in Figure 3 for an initial two or four bioassays. That is, if a CD_{10} is first estimated from two or four bioassays (with males and females of a species counted separately) using the current paradigm of choosing the lowest (significant) CD_{10} , Figure 3 shows what can be expected when that estimate is subsequently updated after 1, 2, 4, or 8 further bioassays. For example, for an initial two bioassays, after one further bioassay there is a 6% chance that the CD_{10} estimate will be reduced by a factor of about 5 or more; after two further bioassays such a change can be expected for about 10% of chemicals; after 4 further bioassays for about 15% of chemicals, and after 8 further bioassays for about 20% of the chemicals tested. Similarly, starting with four bioassays, reductions of a factor of 5 or more in the human CD_{10} estimate can be expected for about 2%, 3.5%, 6%, and 9% of chemicals after a further one, two, four, or eight bioassays.

Examples of the reductions in minimum CD_{10} estimates that occurred over time with increasing numbers of bioassays are shown in Figure 5 for the 31 independent bioassays published for vinyl chloride and the 21 for N-methyl-N'-nitro-N-nitrosoguanidine. For vinyl chloride, the minimum CD_{10} estimate decreased from 1979 to 1988 by a factor of 230, while the minimum CD_{10} estimate for N-methyl-N'-nitro-N-nitrosoguanidine decreased by a factor of 5.0 over the period 1970 to 1993.

To illustrate the contrast with using the current extreme value measure to estimate CD_{10} in humans, Figure 4 shows the effect of instead using median estimate from all available bioassays. In this case the probabilities for a change of a given size in the CD_{10} estimate are lower, and the CD_{10} estimates are as likely to go up as to go down.. For example, with two initial bioassays, there is less than 6% chance for a change by a factor of 5 or more even after eight additional bioassays, and there is equal probability for such a change in either direction.

Discussion

We have demonstrated that the current default paradigm for selecting bioassay results for extrapolation to humans will result in substantial reductions in CD₁₀ estimates in an appreciable fraction of chemicals when they are re-evaluated based on improved data (i.e. additional bioassays); and only reductions (or no change) in CD₁₀ estimates are possible under this paradigm. Such behavior is a substantial disincentive for performing further bioassays, for example to test mechanistic ideas that might serve to improve risk assessment methodologies for individual chemicals.

It is not difficult to devise estimators that reduce the probability for large changes in such circumstances, and symmetrize the direction of potential effects, thus increasing the incentives for additional research. Any such change may call for other changes in the current approaches for extrapolation from laboratory rodents to humans to correctly account for the uncertainty distributions observed in extrapolating between animal species; however, the current approach (based on allometric scaling) is known to be empirically inconsistent with observations, at least in the mean, for extrapolation between mice, rats, and hamsters,^{17,19} so a further examination is certainly overdue. This kind of data and model-driven analysis can lead to action. Especially for that reason, we think that Lester Lave would have liked this analysis.

Acknowledgments

GSO acknowledges support from NIH cooperative agreement grants U54DA021519, UL1 RR024986, RM-08-029, and U54ES017885.

References

1. Lave LB, Omenn GS. Cost-effectiveness of short-term tests for carcinogenicity. *Nature*. 1986; 324(6092):29–34. [PubMed: 3097548]
2. Lave LB, Ennever F, Rosenkranz S, Omenn GS. Information value of the rodent bioassay. *Nature*. 1988; 336(6200):631–633. [PubMed: 3143914]
3. Presidential/Congressional Commission on Risk Assessment and Risk Management. Framework for Environmental Health Risk Management. Vol. 2. United States Government Printing Office; Washington, DC: 1997. <http://www.riskworld.com/Nreports/nr7me001.htm>
4. Purchase IFH. Inter-species comparisons of carcinogenicity. *Br J Cancer*. 1980; 41(3):454–468. [PubMed: 7387835]
5. Gold LS, Bernstein L, Magaw R, Slone TH. 1989 Interspecies extrapolation in carcinogenesis: prediction between rats mice. *Environ Health Perspect*. 1989; 81:211–219. [PubMed: 2759059]
6. Gold LS, Slone TH, Manley NB, Bernstein L. Target organs in chronic bioassays of 533 chemical carcinogens. *Environ Health Perspect*. 1991; 93:233–246. [PubMed: 1773795]
7. Freedman DA, Gold LS, Lin TH. Concordance between rats and mice in bioassays for carcinogenesis. *Regul Toxicol Pharmacol*. 1996; 23(3):225–232. [PubMed: 8812964]
8. Crouch E, Wilson R. Interspecies comparison of carcinogenic potency. *J Toxicol Environ Health*. 1979; 5(6):1095–1118. [PubMed: 529342]
9. Crouch EAC. Uncertainties in interspecies extrapolations of carcinogenicity. *Environ Health Perspect*. 1983; 50:321–327. [PubMed: 6873022]
10. Gaylor DW, Chen JJ. Relative potency of chemical carcinogens in rodents. *Risk Anal*. 1986; 6(3): 283–290. [PubMed: 3602501]

11. Chen JJ, Gaylor DW. Carcinogenic risk assessment: comparison of estimated safe doses for rats mice. *Environ Health Perspect.* 1987; 72:305–309. [PubMed: 3622439]
12. Allen BC, Shipp AM, Crump KS, Kilion B, Hogg ML, Tudor J, Keller B. Investigation of cancer risk assessment methods. (4 parts: Summary and Volumes 1–3). Environmental Protection Agency. 1987 EPA/600/6-87/007a,b,c,d.
13. Allen BC, Crump KS, Shipp AM. Correlation between carcinogenic potency of chemicals in animals and humans. *Risk Anal.* 1988; 8(4):531–544. [PubMed: 3266676]
14. Crump K, Allen B, Shipp A. Choice of dose measure for extrapolating carcinogenic risk from animals to humans: an empirical investigation of 23 chemicals. *Health Physics.* 1989; 57(Sup 1): 387–393. [PubMed: 2606698]
15. U.S. Environmental Protection Agency. Draft report: a cross-species scaling factor for carcinogen risk assessment based on equivalence of $\text{mg/kg}^{3/4}/\text{day}$: Notice. *Federal Register.* 1992; 57:21152–24173.
16. Gaylor DW, Chen JJ, Sheehan DM. Uncertainty in cancer risk estimates. *Risk Anal.* 1993; 13(2): 149–154. [PubMed: 8502788]
17. Crouch EAC. Uncertainty distributions for cancer potency factors: laboratory animal carcinogenicity bioassays and interspecies extrapolation. *Human Ecol Risk Assess.* 1996; 2(1): 103–129.
18. Gold, LS.; Ames, BN.; Bernstein, L.; Blumenthal, M.; Chow, K.; Da Costa, M.; de Veciana, M.; Eisenberg, S.; Garfinkel, GB.; Haggin, T.; Havender, WR.; Hooper, NK.; Levinson, R.; Lopipero, P.; Magaw, R.; Manley, NB.; MacLeod, PM.; Peto, R.; Pike, MC.; Rohrbach, L.; Sawyer, CB.; Slone, TH.; Smith, M.; Stern, BR.; Wong, M. The Carcinogenic Potency Database (CPDB). 2008. <http://potency.berkeley.edu/>
19. Crouch, EAC. Cancer risk assessment: more uncertain than we thought. In: Hu, C-H.; Stedefort, T., editors. *Cancer Risk Assessment*. John Wiley & Sons, Inc; Hoboken, New Jersey: 2010.
20. Guidelines for Carcinogen Risk Assessment. Risk Assessment Forum, U.S. Environmental Protection Agency; Washington, DC: 2005. EPA/630/P-03/001F
21. Faustman, EM.; Omenn, GS. Risk assessment. In: Klaassen, CD., editor. *Casarett & Doull's Toxicology: The Basic Science of Poisons*. 8. New York: McGraw-Hill; 2012. in press

Figure 1.

Examples of within-species distributions of maximum likelihood CD_{10} estimates from 44 distinct bioassays of 2-acetylaminofluorene (mouse), 31 of vinyl chloride (rat), and 21 of DDT (mouse) (data from Gold et al.¹⁸ as analyzed by Crouch¹⁹). Error bars are approximate 1 SD. Plotted are $\ln(CD_{10})$ estimates in rank order versus the normal score. These empirical distributions are indistinguishable from lognormal distributions ($p > 0.26$, Shapiro-Wilk test), which would plot as approximate straight lines on this figure.

Figure 2. Distribution across 331 chemicals of the logarithm of the ratio of rat and mouse median CD_{10} estimates (data from Gold et al.¹⁸ as analyzed by Crouch¹⁹). Plotted are rank ordered differences in mean $\ln(CD_{10})$ estimates in rat and mouse versus normal score. Error bars are approximate 1 SD. Red line is maximum likelihood fit to a lognormal distribution.

Figure 3.

Cumulative probabilities for the relative change in estimated human CD_{10} after 1, 2, 4, or 8 further bioassays, based on two (top pane) or four (bottom pane) initial bioassays, using the current paradigm.

Figure 4. Cumulative probabilities for the relative change in estimated human CD_{10} after 1, 2, 4, or 8 further bioassays, based on two (top pane) or four (bottom pane) initial bioassays, using a median measure.

Figure 5.

CD_{10} estimates from published independent bioassays, and the observed trajectory of the minimum CD_{10} estimate, versus publication year, for vinyl chloride (31 bioassays) and N-methyl-N'-nitro-N-nitrosoguanidine (21 bioassays).

Parameters of the distributions across chemicals of logarithms of ratios of median (within species, across bioassays) CD_{10} estimates (in mg·kg/d).

Table 1

Parameter of distribution	$\ln(CD_{10}[Rat]) - \ln(CD_{10}[Mouse])$		$\ln(CD_{10}[Mouse]) - \ln(CD_{10}[Hamster])$		$\ln(CD_{10}[Hamster]) - \ln(CD_{10}[Rat])$	
	Maximum likelihood estimate	SD	Maximum likelihood estimate	SD	Maximum likelihood estimate	SD
Mean	-0.24	0.11	-1.06	0.42	0.98	0.32
Standard deviation	1.83	0.08	1.94	0.31	1.74	0.24

Table 2

Parameters of lognormal distributions describing the standard deviations of the natural logarithms of the within-species, across-bioassay, CD₁₀ distributions.

	Rat	Mouse	Hamster
Median	0.49	0.36	0.23
Standard deviation of natural logarithm	0.79	0.88	0.68