Journal of Hazardous Materials, 14 (1987) 283—292
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

THE USE OF ANIMAL MODELS FOR PREDICTING THE POTENTIAL CHRONIC EFFECTS OF HIGHLY TOXIC CHEMICALS

C.G. ROUSSEAUX

Department of Veterinary Pathology, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, Saskatchewan S7N 0W0 (Canada)

(Received April 21, 1986; accepted in revised form July 18, 1986)

Summary

The acute effects of highly toxic chemicals in man and animals are often blatantly obvious. However, the data obtained from experimental investigation of the chronic effects of toxic chemicals often correlate poorly with epidemiological findings following exposure of the population to these chemicals in the field. The reasons for this discrepancy between experimental and epidemiological observations are numerous and complex. The areas of erroneous predictions resulting from errors in data extrapolation, inaccurate statistical methods, environmental variables and a lack of standard operating procedures may account for some of the lack of correlation between experimental findings and the in-field situation. Epidemiological studies may not give a true assessment of the impact of a chemical on the population as a whole.

Introduction

Chemical accidents, such as the tragic events in Bhopal, India [1], can produce acute, disastrous and fatal effects. The media-heightened public awareness of the acute effects of these chemicals (spectacular nature of numerous deaths in a disaster such as this) cannot be avoided [2, 3]. However, little is said concerning the more long-term or chronic effects of highly toxic chemicals. Such effects may include the production of cancerous growths (carcinogenesis), abnormal fetal development (teratogenesis), abnormal reproduction and chronic, irreversible pathological changes. Epidemiological investigations into the chronic effects of highly toxic chemicals accidentally released into the biosphere are limited, and are often in conflict with experimental data. At present, data concerning the long-term effects of methyl isocyanate, released in the Bhopal incident, are limited to subjective reports of persistent "respiratory problems" [1]. However, a great deal of interest has been shown in the long-term effects of this highly toxic chemical [4]. Because the data from the Bhopal incident are so incomplete, laboratory investigations of other chemicals will be compared with the epidemiological data available for accidental

0304-3894/87/$03.50

release and, in some instances, pharmacological use of these chemicals to illustrate how and why discrepancies occur among these data.

The investigation of the chronic effects of many toxic chemicals have often been fraught with poor experimental design, data analysis and interpretation of results. Frequently, epidemiological data obtained from the human population fail to correlate with experimental results indicating that either false-positive of false-negative risk assessments have been made using data obtained in the test species used. For example, in the Seveso incident in Italy, where a large amount of 2,3,7,8-tetradichlorodibenzo-*p*-dioxin (TCDD) or dioxin, was spilled into the atmosphere, epidemiological data failed to correlate with results obtained from experiments in the laboratory [5]. Considerable testing in rodents and largomorphs revealed that dioxin was, in these species, a potent teratogen [6, 7], a carcinogenic promotor [8], and a hepatotoxin [9]. TCDD has also been associated with reproductive toxicity [10], platelet depression [11], immune disturbances [12], chloracne [5] and extreme toxicity in low concentration [13]. During the Seveso accident, large amounts of dioxin were released into the atmosphere, which were sufficient to produce a fall-out of 15 mg of dioxin per kilogram of grass 900 meters from the factory. The only clinical finding in humans exposed to this chemical was chloracne. All other chronic effects were determined to be negligible, and were not statistically greater than in other members of the population not exposed to that chemical [5]. In this instance it appears that false-positive experimental results did not produce adequate, or accurate information concerning the effects, or risk of effects to the exposed local population.

The converse, where false-negative experimental results failed to predict the drastic effects of a "non-toxic" chemical on the exposed population, has also been shown to be true. Epidemiological data have been used to demonstrate that thalidomide was a potent teratogen in humans [14], whereas experimentally it appeared to be benign in rodents [15]. Similarly, it was only recently that the impact of diethylstilbestrol use has been seen in young women exposed *in utero*, in the form of neoplasia of the reproductive tract [16]. This drug had previously passed all licensing requirements [17], that recommended it to be "safe" for use in humans.

The reason why results from cellular and animal experiments, often do not give an accurate assessment of risk for the human population at large, is complex. Most erroneous predictions occur because of extrapolation from cellular or animal models to man, inaccurate interpretations from inappropriate use of statistical methods when evaluating data, lack of standard operating procedures, and unforeseen environmental effects. Finally, sample size and sample bias in epidemiological studies may effect evaluation of the effect of a chemical on the whole exposed population.

Data extrapolation

Extrapolating data from animals to man, or experimental models to

man, is fraught with difficulties. Firstly, obvious genetic and metabolic differences between the model and man may account for some differences seen between the epidemiological and laboratory observations. *In vitro* and *in vivo* tests are often used to predict the in-field situation. The Ames test, where *Salmonella* bacteria are used to determine the mutagenicity of a compound [18], may be a reasonable indication of mutagenicity in bacteria; however, it may fail to predict mutagenicity in large mammalian organisms [19]. The old adage of "the difference between mice and men" is particularly true when laboratory rodents, and non-rodent mammals, are used to test carcinogenicity, long term toxicity and teratogenicity of compounds. These differences occur among species and among strains of the same species [20, 21]. In other words, different strains that are exposed to the same chemical can quite often differ statistically among each other when similar end points are measured. Data obtained from these experiments are used for licensing purposes [22, 23]. Using *in vitro* and *in vivo* models for testing compounds to predict possible detrimental effects to humans is necessary. But unfortunately, blind acceptance of the experimental results, with respect to risk assessment in man is sometimes made, without evaluating the possibility of false-negative or false-positive results.

Teratological studies have their own particular problems. Firstly, every compound is apparently teratogenic [24]. It is only a matter of finding the narrow window in the dose scale at which defective development occurs. Secondly, the rodent, and sometimes rabbit models, can produce numerous, and sometimes statistically significant numbers of non-specific terata when given concentrations of a compound that produces maternal toxicity [25]. These two observations mean that compounds are often labelled as teratogens when, in actual fact, they are not [26]. To be a teratogen a compound should either produce defective development at a level that is not toxic to the mother, or produce specific repeatable defects at levels that are toxic to the mother [26].

Similarly problems have been experienced with carcinogenicity studies when extrapolations to man were made. Most carcinogencity studies demand almost lifelong administration of the compound [27]. As a number of tumors are age related and have different clinical meanings with respect to the rate of growth, occurrence of metastases, and subsequent death of the individual, sorting out significant biological or statistical differences is often difficult [28]. In fact, there is a 25—50% chance of having a false-positive results for a compound if the standard statistical formulation of tests for the null hypothesis are used [27].

Classification of masses into hyperplasia, benign tumor or malignancy is based upon an opinion of a pathologist [29]. Situations arise where this judgement varies widely from pathologist to pathologist [30]. Therefore, it is essential that one pathologist examines all the study to keep operator bias to a minimum [31]. In fact, all slides should be read by

one pathologist without prior knowledge from which treatment group the slides came. Attempts have been made to grade the severity of pathological changes as mild, moderate or severe so that comparisons can be made [31]. However, subjective ranking on a 5 point scale that has been recently used may be superior [32]. This ranking provides the means for application of nonparametric statistical analysis. Variation among pathologists in both the rank of severity and, the spread of use of the scale, may result in invalid statistical interpretations. Data generated from studies in which no pathologist was used are open to even more variation, and are somewhat questionable.

Statistical methods

Statistical analysis of data may also account for some of the differences between experimental data and epidemiological data gained from human exposure to chemicals. In the past, claims to have an effect where no statistical evaluation of data had been undertaken were numerous, particularly in the field of teratology (an example is given in Ref. [33]). Sometimes percentages and means were considered sufficient "proof" but are in actual fact only descreptive statistics [34]. Without some means of evaluating the variation or spread of the data, probabilistic testing of the Neymen-Pearson or "null" hypothesis is impossible [35].

Often the wrong statistical test has been applied to data to obtain the required probability (P) value. Grading of lesions, ranking the severity of defects, etc., constitute ordinal data and not interval data such as organ weights, body measurements, etc. Various non-parametric statistical analyses are available for ordinal and nominal data [36]. Failure to use the correct procedure to test for differences in the Neyman-Pearson hypothesis may have resulted in erroneous probability values. This problem has been further exacerbated by abuse of packaged computer programs. Since their introduction in the 1960s, packaged computer programs have freed researchers from much of the pain-staking computational labour of statistical analysis. Although many meaningful projects could not have been attempted without these packages, the very ease and use of these packages has led to their abuse. In particular, performing all possible T- tests among numerous groups, calculating many correlations, and circling the ones with probability values less than 0.05, and performing step-wise regressions, all can easily lead to spurious statistical conclusions [35]. Even when the correct test is applied to the corresponding data, readers and writers often forget that a probability value of 0.05 means that there is a 1 in 20 chance that the difference among groups may be due to chance alone.

Spurious results can be produced by not correctly defining the experimental unit. Usually one assumes that the experimental unit is the test animal or the *in vitro* test unit but often the results are not expressed as a function of this experimental unit. This mistake is common in the teratology studies,

where results can often be expressed as defects per fetus in a treatment group. The mistake here is that the dam should have been used as the experimental unit [37] and the defect should be expressed as defects per fetus per dam [37]. Similarly, carcinogenicity studies should express findings per individual rather than so many tumors per treatment group. In the case of carcinogenicity studies it has been proposed that rather than Neyman-Pearson hypothesis testing, more complicated mathematical models be used to eliminate some of the false-positive findings that arise from incorrect groupings [27].

Standard operating procedures

Standard operating procedures and good laboratory practice guidelines are used to try and eliminate some of the day-to-day variation among technicians, and stabilize environmental variables so that comparison of the test article can be made with other environmental variables being held constant [38]. Failure to comply with standard operating procedures, or to use standard operating procedures, may result in data that are not suitable for extrapolation or to form the basis for risk assessment of the population at large.

Environmental variables that exist during the experiment, but not in the field, may account for some differences that have been observed between experimenter's results in the field situation. Rodents, and other experimental animals, can come into contact with compounds in their local environment, which may effect the toxicity of a test compound. Particularly of note is bedding material used in the cages of experimental animals. Here, pine chips, and other bedding material that contain hydrocarbons, may induce biotransforming enzymes, in particular the cytochrome P-450 oxidase group [39]. Such induction of biotransforming enzymes may change the animal's ability to biotransform toxic compounds, to either a more toxic compound or a harmless metabolite [40]. The effects of this biotransformation is not always taken into account by using negative controls.

Rodents, in particular, respond differently to stress than humans. Background high-frequency noise [41], handling and any other maternal stress may affect reproduction and produce increased numbers of defective fetuses mediated through change in the dam [25, 26], whereas these observations have not been made in humans. Diurnal rhythms and seasonal factors [42, 43], sex [44], nutritional status [45], age and body weight [46] may confuse the issue further. All of these factors have been associated with changes in metabolic activities [46] and hence the toxicity produced by xenobiotics. Standard operating procedures attempt to control these variables. Even though these variables are controlled within the experiment, they may account for some of the differences seen between the experimental and epidemiological data.

Methods of chemical administration

Routes of administration are also important with respect to absorption, distribution, and biotransformation of a compound [47]. Often toxic chemicals are administered to a rodent or other laboratory species through a rather unusual route such as intravenous or intraperitoneal. This may yield valuable information for pharmacologists, concerned with some aspects of biotransformation, but does not mimic the exposure seen in the "real-life" situation. Experiments using such routes of application may be poor predictors of the action of chemicals in individuals exposed by oral, respiratory or dermal routes.

The concentration of a chemical administered during experiments when compared with accidential exposure may differ. Often large, but sub-lethal, doses of toxic compounds are administered over a prolonged period in order to produce positive results. It may be necessary to produce experimental toxicity for prolonged periods of time to see what toxicological effects result, but this does not necessarily equate with what we see in the field. Often, in order to be able to produce long-term carcinogenic, teratogenic or reproductive effects experimentally [25, 26], a narrow-dose range must be utilized for a protracted period. Unfortunately, data gained from such experimentation [25, 26], although necessary for evaluation of the toxicity of a compound, are used to produce a black and white verdict concerning the compound's teratogenicity, carcinogenicity and long-term pathological effects.

Non-scientific interference

To some extent risk assessment, based upon data obtained from experimental models is limited by public opinion and the media. Usually, when the carcinogenicity of a compound is even preliminarily suspected it can be legislated to be a carcinogen through media and public pressures [48], thus cementing a positive result in place, be it true or false. The position of animal rights activists and the growing unwillingness of the population to allow experimentation on animals that serve as good models for human exposure, will lead to more *in vitro* testing, and subsequent possible increased false-negative and false-positive results.

Epidemiological studies

Finally, epidemiological evaluation may be the reason for failure of correlation between experimental data and the in-field situation. Grouping and sampling of the population for a retrospective or prospective epidemiological study can result in work-up bias, test-review bias and incorporation bias [49]. This may result in false-positive or negative-results, with respect to the real effect on the population. Similarly, by selecting too few in-

dividuals from the population, false-negative findings are common [50]. In both cases the experimental design, methodological approach, and statistical analysis of an experiment to evaluate risk to the population following chemical exposure may have been correct, but the epidemiological or retrospective study in these situations has failed to predict the real situation throughout the population.

Conclusions

In summary, laboratory data, obtained from animal models used for predicting the potential chronic effects of highly toxic chemicals, are extremely useful in giving qualitative indication of risk to the population following accidental exposure to these toxic chemicals. This is providing that correct interpretations have been made from the data following correct data analysis. No accurate quantitative assessments of risk can be made from data generated from the laboratory, because of false-positive and false-negative results. Laboratory results are often misleading, because of different subjective interpretations, erroneous extrapolations, poor experimental technique, abnormal groupings or incorrect evaluation of data. Conversely, epidemiological results may be inacurrate because of bias in the method of sampling and size of the sample group.

The net effect of the differences between experimental and epidemiological data is to separate individuals into two groups of thought: those who don't believe in subtle, but chronic effects of minute quantities of toxic chemicals; and those who believe that any toxic chemical always produces devastating long term biological effects. With the case in hand, highly toxic chemicals are likely to produce acute toxicity and probably death, as in the Bhopal case. However, the long-term effects on the individual following one exposure of these highly toxic chemicals at a dose at which overt clinical disease does not occur, are likely to be minimal.

Hopefully, the use of animal models for predicting the potential chronic effect of highly toxic chemicals will continue with increased awareness of what assumptions are being made when evaluating the risk to man.

Acknowledgement

I wish to thank Mrs. J. Brown for typing this manuscript.

References

1 J.M. Dave, The Bhopal methyl isocyanate (MIC) incident: An overview, In H.B. Schiefer (Ed.), Highly Toxic Chemicals: Detection and Protection Methods, Proc. Int. Symposium, Saskatoon, Sask., Canada, 1985, University of Saskatchewan Printing Services, pp. 1—39.

2 P. Stoler, Frightening findings at Bhopal: Union Carbide and India begin to uncover what happened, Time, February 18, 1985, p. 56.

3 M. Heylin (Ed.), Bhopal: The continuing story, Chem. Eng. News, 63 (1985) 3—40.
4 W. Lepkowski, Questions persist about cyanide poisoning in Bhopal disaster, Chem. Eng. News, 63 (1985) 42—43.
5 G. Reggiani, Anatomy of a TCDD spill: The Seveso accident, In J. Saxena (Ed.), Hazard Assessment of Chemicals: Current Developments, Vol. 2, Academic Press, Toronto, 1983, pp. 269—342.
6 F.A. Smith, B.A. Schwetz and K.D. Nitschke, Teratogenicity of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in CF-1-mice, Toxicol. Appl. Pharmacol., 38 (1976) 517—523.
7 E. Giavani, M. Piati and C. Vismra, Rabbit teratology study with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, Environ. Res., 27 (1982) 74—78.
8 H.C. Pitot, T. Goldsmith, H.A. Campbell and A. Poland, Promotion by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin of hepatocarcinogenesis from diethylnitrosamine, Cancer Res., 40 (1980) 3616—3020.
9 A. Poland, W.F. Greenlee and A.S. Kende, Studies on the mechanisms of action of the chlorinated-*p*-dioxins and related compounds, Ann. N.Y. Acad. Sci., 320 (1979) 214—230.
10 F.J. Murray, F.A. Smith, K.D. Nitschke, C.G. Humiston, R.J. Kociba and B.A. Schwertz, Three generation reproduction study of rats given 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in the diet, Toxicol. Appl. Pharmacol., 40 (1979) 241—252.
11 J.G. Vos, J.A. Moore and J.G. Zinkl, Toxicity of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in C57B1/6 mice, Toxicol, Appl. Pharmacol., 29 (1974) 229—241.
12 J.G. Vos and J.A. Moore, Suppression of cellular immunity in rats and mice by maternal treatment with 2,3,7,8-tetradichlorodibenzo-*p*-dioxin, Int. Arch. Allergy Appl. Immunol., 47 (1974) 777—794.
13 B.A. Schwetz, J.M. Norris, G.L. Sparschu, V.K. Rowe, P.J. Gehring, J.L. Emerson and C.G. Gerbidg, Toxicology of chlorinated dibenzo-*p*-dioxins, Env. Health Persp., 5 (1973) 87—99.
14 W.G. McBride, Thalidomide and congenital abnormalities, Lancet, 2 (1961) 1358.
15 J. Warkany, Congenital Malformations, Notes and Comments, Year Book Publishers, Chigaco, 1971, pp. 102—125.
16 S.J. Robboy, W.R. Welch and J. Prat, Intrauterine exposure to diethylstilbestrol, In R.H. Riddell (Ed.), Pathology of Drug-Induced and Toxic Diseases, Churchill-Livingstone, New York, 1982, pp. 325—340.
17 K.L. Noller and C.B. Fisch, Diethylstilbestrol usage: its interesting past, important present, and questionable future, Med. Clin. North Amer., 58 (1974) 739—810.
18 B.N. Ames, J. McCann and E. Yamasaki, Methods for detecting carcinogens and mutagens with the *Salmonella* mammalian microsome mutagenicity test, Mutat. Res., 31 (1975) 347—364.
19 E.J. Calabrese, Principles of animal extrapolation, In R.L. Metcalf and W. Stumin (Eds.), A Volume in Environmental Science and Technology, Wiley-Interscience, 1983, pp. 1—599.
20 H. Frohberg, On the problems of transposing experimental teratogenicity trials to man, In A. Spiegel (Ed.), The Laboratory Animal in Drug Testing, 5th I.C.L.A. Symposium, Hanover, 1972, Gustav Fischer Verlag, Stuttgart, 1973, pp. 85—97.
21 H.B. Schiefer, Use and evaluation of test animals, In J. Beare-Rogers (Ed.), Methods for Nutritional Assessment of Fats, American Oil Chemistry Society, 1985, pp. 17—43.
22 J.A. Zapp, Jr., Extrapolation of animal studies to the human situation, J. Toxicol. Environ. Health, 2 (1977) 1425—1433.
23 D.B. Clayson, D. Krewski and I. Munro, Toxicological Risk Assessment, Vol. 1, Biological and Statistical Criteria, CRC Press, 1985.
24 J.G. Wilson, Environmental Birth Defects, Academic Press, New York, 1973.

25 K.S. Khera, Maternal toxicity: A possible factor in fetal malformations in mice, Teratology, 29 (1984) 411—416.
26 K.S. Khera, Maternal toxicity: A possible etiological factor in embryo—fetal deaths and fetal malformations of rodent—rabbit species, Teratology, 31 (1985) 129—153.
27 D.S. Salsburg, Use of statistics when examining lifetime studies in rodents to detect carcinogenicity, J. Toxicol. Environ. Health, 3 (1977) 611—628.
28 R. Doll, The age distribution of cancer: Implications for models of carcinogenesis, J. Roy, Stat. Soc. A, 134 (1971) 133—136.
29 K.S. Crump, D.G. Hoel, C.H. Langley and R. Peto, Fundamental carcinogenic processes and their implications for low dose risk assessment, Cancer Res., 36 (1976) 2973—2979.
30 S. Takayama, Variation of histological diagnosis of mouse liver tumors by pathologists, In W.H. Butler and P.M. Newberne (Eds.), Mouse Hepatic Neoplasia, Elsevier, Amsterdam, 1975, pp. 183—187.
31 J.F. Robens, J.J. Joiner and R.L. Schueler, Methods in testing for carcinogenicity, In A.W. Hayes (Ed.), Principles and Methods of Toxicology, Raven Press, New York, 1982, pp. 79—105.
32 M.A. Hayes, J.E.C. Bellamy and H.B. Schiefer, Subacute toxicity of dietary T-2 toxin in mice: Morphological and hematological effects. Can. J. Comp. Med., 44 (1986) 203—218.
33 L.S. Hurley, E. Wooten, G.J. Everson and C.W. Asling, Anomalous development of ossification in the inner ear of offspring of manganese deficient rats, J. Nutrit., 71 (1960) 15—18.
34 D.R. Cox, Statistical significance tests, Br. J. Clin. Pharmacol., 14 (1982) 325—331.
35 C.F. Hofacker, Abuse of statistical packages: The case of the general linear model, Amer. J. Physiol., 245 (Regulatory Integrative Comp. Physiol., 14) (1983) R299—R302.
36 J.D. Emerson and G.A. Colditz, Use of statistical analysis in The New England Journal of Medicine, New Eng. J. Med., 309 (1983) 709—713.
37 Middle Atlantic Reproduction and Teratology Association Workshop on Statistical Design and Analysis of the Segment II Teratological Study (1975).
38 The selection of doses in chronic toxicity/carcinogenicity studies, In H.C. Grice, Current Issues in Toxicology, Springer-Verlag, New York, 1984, pp. 9—41.
39 R.A. Neal, Metabolism of toxic substances, In J. Doull, C.D. Klaassen and M.O. Amdur (Eds.), Casarett and Doull's Toxicology: The Basic Science of Poisons, Collier Macmillan, Toronto, 1980, pp. 56—69.
40 M.R. Juchau and S.T. Chao, Drug metabolism by the human fetus, Clin. Pharmacokinet., 5 (1980) 320—339.
41 P.S. Nawrot, R.O. Cook and R.E. Staples, Embryotoxicity of various noise stimuli in the mouse, Teratology, 22 (1980) 279—289.
42 I. Chahoud, Measurement of circadian rhythm in the motility of weaned rodents, In D. Neubert, H.J. Merker and T.E. Kwasigroch (Eds.), Methods in Prenatal Toxicology, Thieme, Stuttgart, 1977, pp. 103—112.
43 J.A. Romero, Influence of diurnal cycles on biochemical parameters of drug sensitivity: The pineal gland as a model, Fed. Proc., 35 (1976) 1157—1161.
44 A.P. Otis and H.L. Foster, Management and design of breeding facilities, In H.L. Foster, J.D. Small and J.G. Fox (Eds.), Mouse in Biomedical Research, Vol. III, Academic Press, New York, 1983, pp. 18—35.
45 E.M. Boyd, Predictive Toximetrics, Science Technology, Bristol, 1972.
46 J. Doull, Factors influencing toxicology, In J. Doull, C.D. Klaassen and M.O. Amdur (Eds.), Casarett and Doull's Toxicology: The Basic Science of Poisons, Collier Macmillan, Toronto, 1980, pp. 70—83.
47 C.D. Klaassen, Absorption, distribution and excretion of toxicants, In J. Doull, C.D. Klaassen and M.O. Amdur (Eds.), Casarrett and Doull's Toxicoloy: The Basic Science of Poisons, Collier Macmillan, Toronto, 1980, pp. 28—55.

48 E. Efron, The Apocalyptics: Cancer and the Big Lie: How Environmental Politics Control What We Know About Cancer, Simon and Schuster, New York, 1984.
49 D.F. Ransohoff and A.R. Feinstein, Problems of spectrum and bias in evaluating the efficacy of diagnostic tests, New Engl. J. Med., 299 (1978) 926—929.
50 J.A. Frieman, T.C. Chalmers, H. Smith and R.R. Kuebler, The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial, New Engl. J. Med., 299 (1978) 690—694.