

The Excel Format of the Carcinogenic Potency Database

Lois Swirsky Gold, Thomas H. Slone, and Neela B. Manley

The Excel format is derived from the Carcinogenic Potency Database (CPDB) of Gold *et al.* that has been published in several earlier papers. This Excel format includes the same information as the plot version, and therefore the methods and guide to the plot provide details on what is summarized here.

The CPDB is a single standardized resource of the results of 45 years of chronic, long-term carcinogenesis bioassays. The experiments vary widely in design, histopathological examination and nomenclature, and in the published authors' choices of what information to publish in their papers. Data are included from 6153 experiments reported in the general literature and in the in Technical Reports of the National Cancer Institute/National Toxicology Program (NCI/NTP). Information is given in the CPDB on strain, sex, route of compound administration, target organ, histopathology, author's opinion about carcinogenicity, and reference to the published paper, as well as quantitative data on statistical significance, tumor incidence, dose-response curve shape, length of experiment, duration of dosing, and dose-rate.

The files on this Web site for the Excel format include (A) documentation of methods, field descriptions, and linking instructions; (B) Excel files; and (C) ancillary files of appendices.

I. File Organization

The Carcinogenic Potency Database (CPDB) contains the following files:

File Name	Description
<u>Documentation</u>	
Excel_CPDB.pdf (this file)	This file of brief documentation describes 1) Methods: inclusion criteria for experiments, literature review, and rules for tissue and tumor inclusion 2) data and linking instructions for the Excel files 3) Structure and fields of the Excel files
Documentation_for_Excel.pdf	A detailed description of the fields in CPDB datasets (rich text format)
Methods.pdf	The methods of the CPDB (rich text format)
CPDB.pdf	A guide to the CPDB (rich text format)
<u>CPDB data</u>	
cpdb.ncintp.xls	All data for NCI/NTP except doses and incidence
cpdb.ncntdose.xls	Dose and incidence data for NCI/NTP
cpdb.lit.xls	All data for literature except: 1) doses and incidence (cpdb.litdose.xls) 2) supplementary data reported in cpdb.supp.xls and cpdb.suppvary.xls
cpdb.litdose.xls	Dose and incidence data for literature
cpdb.chemname.xls	Chemical names, three-letter identification codes and CAS numbers
cpdb.cit.xls	Brief citation to published paper in literature
cpdb.species.xls	Species code definitions
cpdb.route.xls	Route code definitions
cpdb.strain.xls	Strain code definitions
cpdb.tissue.xls	Tissue code definitions
cpdb.tumor.xls	Tumor histopathology code definitions
cpdb.journal.xls	Journal code definitions

Appendices

Bibliography.pdf	A bibliography of literature citations (rich text format)
Note codes.pdf	A list of note code definitions (rich text format)
Monkey documentation.pdf	Inclusion and analysis rules for NCI monkey studies

II. METHODS: SUMMARY

In the CPDB, an experiment is defined as the control and dose groups under a set of bioassay conditions that includes a single species, sex, strain of mammal administered a test agent by a given route for a given exposure and experiment length. The sources of experimental results are the NCI/NTP Technical Reports and the general published literature. Further details on methods are in Methods.pdf and CPDB.pdf.

II. A. Inclusion Criteria

1. Animals on test were inbred or outbred mammals;
2. Administration was begun early in life (100 days of age or less for rats, mice, and hamsters); in utero exposure is not included.
3. Route of administration was diet, water, gavage, inhalation, intravenous or intraperitoneal injection.
4. Test agent was a single chemical or commercial chemical mixture; particulate matter is not included.
5. Test agent was administered alone, rather than in combination with other chemicals;
6. Exposure was chronic, with not more than 7 days between administrations;
7. Duration of exposure was at least one-fourth the standard lifespan for that species; (rodent standard lifespan is 2 years).
8. Duration of experiment was at least half the standard lifespan for that species;
9. Research design included a concurrent control group;
10. Research design included at least 5 animals per group;
11. Surgical intervention was not performed;
12. Microscopic histopathology data were reported for the number of animals with tumors rather than the total number of tumors;
13. Results reported were original data, rather than secondary analyses of experiments already reported by other authors.
14. For experiments with serial sacrifices, data are reported separately for each sacrifice and animals were preselected for sacrifice.
15. For long-term studies in monkeys conducted by NCI, some inclusion rules have been relaxed. (See Monkey documentation.pdf).

II. B. Bioassay Search

All NCI/NTP carcinogenesis bioassays that meet the inclusion criteria and are published in Technical Reports are included in the CPDB.

A search of the published literature has been conducted for all experiments which meet the standard criteria, whether or not the authors considered the test agent related to tumor induction.

The literature search covered: the *Survey of Compounds Which Have Been Tested for Carcinogenic Activity* (formerly PHS 149), Medline, Current Contents, the Monographs on chemical carcinogens prepared by the International Agency for Research on Cancer, the FDA database of food additives, and the Japanese Science and Technology Database.

Separate journal searches: *British Journal of Cancer*, *Cancer Letters*, *Cancer Research*, *Carcinogenesis*, *Chemosphere*, *Environmental Health Perspectives*, *European Journal of Cancer and Clinical Oncology*, *Food and Chemical Toxicology*, *Fundamental and Applied Toxicology*, *International Journal of Cancer*, *Japanese Journal of Cancer Research*, *Journal of Cancer Research and Clinical Oncology*, *Journal of Environmental*

Pathology and Toxicology, Journal of Toxicology and Environmental Health, Journal of the National Cancer Institute, Regulatory Toxicology and Pharmacology, Toxicology, and Toxicology and Applied Pharmacology.

II. C. Selection of Tissue and Tumor Types for Inclusion

A tissue-tumor combination is reported in the CPDB for each category of neoplasm, benign or malignant, which an author evaluated as treatment-related, regardless of the statistical or biological basis for the evaluation. (Hyperplasia and non-neoplastic lesions are not included in the database.) Results indicating a reduction in tumor incidence compared to controls are not included. The nomenclature for tissues and tumors, as defined by codes in the CPDB, reflects the terminology used in the NCI/NTP Report or by the authors of a published paper. See the file “Documentation_for_Excel.pdf” and the datasets cpdb.tissue.xls and cpdb.tumor.xls.

To facilitate comparisons across experiments and chemicals, the CPDB includes three additional categories of tissue-tumor combinations: “mandatory sites,” “all tumor-bearing animals” and “statistical sites.” For NCI/NTP there is a prescribed set of mandatory sites (see below). In the general literature, whenever there is adequate documentation in the published paper, the same tissues as in the NCI/NTP mandatory sites are included in CPDB.

II. D. NCI/NTP Bioassays

The selection rules for tissue and tumor combinations from NCI/NTP Technical Reports are as follows:

1. All tissues are included when the NTP opinion was “clear”, “some” or “equivocal” evidence of carcinogenic activity or the NCI opinion was “carcinogenic” or “associated” with carcinogenicity.
2. The category “all-tumor bearing animals,” excluding interstitial-cell tumors of the testis for Fischer 344 rats.
3. Mandatory sites: Liver in rats and mice: hepatocellular adenoma, hepatocellular carcinoma, or neoplastic nodule. From more recent Technical Reports, hepatocellular adenoma, hepatocellular carcinoma, or hepatoblastoma. Also in mice, lung alveolar bronchiolar adenoma and alveolar bronchiolar carcinoma.
4. Statistical sites: In the absence of an opinion by NCI/NTP, a tissue-tumor combination is eligible to be included in the CPDB if it is reported in the statistical tables in the Technical Reports. The choice of which statistical tests and which sites NCI or NTP has chosen to report has changed over time. For a statistical site to be eligible for inclusion in the CPDB, the statistical tables must report a positive (not negative) and statistically significant result ($p < 0.1$) for at least one of the statistical tests. From among these sites, the CPDB includes only those for which the two-tailed p -value by lifetable analysis is less than 0.05 for the test that the slope of the dose response is different from zero (likelihood ratio statistic). See section II.G., “Statistical Significance and Dose Response.”
5. Berkeley Mixes: Because the computer files from the individual animal pathology tables of the NCI/NTP give pathology results on each animal, it has been possible to report results for a composite category of all tumors which the NCI/NTP evaluated as treatment-related. These are described in the file “Documentation_for_Excel.pdf” for the field mixberk. It has generally not been possible to formulate such a composite category for other experiments in the database, with the exception of a group of studies on aromatic amines, bioassays in monkeys, and a few other experiments in the CPDB.

For several recent bioassays in which NTP suspected a potential for chemically induced renal tubule neoplasms, multiple additional sections were taken of the kidneys. CPDB reports results for the standard protocol and additionally for the combined standard and step sections. See the file “Documentation_for_Excel.pdf” for the field step.

II. E. Bioassays in the Published Literature

The great diversity in the published bioassay literature over time necessitates many decisions about inclusion of experiments and of particular tissue-tumor combinations. We have been able to augment the

results in published papers through personal communication with hundreds of authors of published papers, particularly for experiments conducted recently, e.g., to clarify opinions as to carcinogenicity at particular sites, to obtain additional tumor data or separate data for experiments with interim sacrifices.

In the general literature, results in the CPDB are often given separately for benign and malignant tumors at a given site because it is not known whether the pathologist scored animals only for the most malignant tumor. For example, it may be impossible to know whether the proportion of animals with adenomas represents only those animals that did not have carcinomas. Therefore, no mix could be created of benign and malignant tumors at a given site.

For each experiment in the general literature, the selection rules for tissues and tumors are as follows:

1. Tissue-tumor combinations that the published author evaluated as induced by the test agent.
2. The category “all tumor-bearing animals”
3. Mandatory sites: Liver for rats, mice, and hamsters; lung in mice and hamsters.

Authors of papers in the general literature rarely indicate which animals were diagnosed as having more than one tumor. Therefore, it has been generally impossible to combine data on various tumors within a single tissue, or to specify particular tumor types which should be included in mandatory sites. To attempt to combine incidence data would risk multiple counting of animals.

4. Statistical sites: In the interest of completeness, results are included for tissues and tumors in the literature which the authors did not consider treatment-related, but for which the increase in tumor incidence is statistically significant (χ^2 one-sided $p < 0.05$) for any dose-group compared to controls.

In the general literature, tissue-tumor combinations can be included that are not statistically significant (χ^2 $p < 0.05$) if there were no results in the paper for mandatory sites, and it was therefore necessary to choose some site so that the experiment could be included in the database. Additionally, some tissue-tumor combinations that are not statistically significant are included from the literature for toxicological interest, e.g. if the site is rarely a site of tumorigenesis, and the experiment was conducted to determine if there was an effect at that site.

II. F. Estimation of Average Daily Dose Level

Because a variety of routes of administration, dosing schedules, species, strains, and sexes are used in carcinogenesis bioassays, some standardization of dose is required. Our convention is to determine for each dose group in an experiment the daily dose-rate in mg/kg of body weight averaged over the duration of the experiment. To convert ppm or percent in food, water, or air, into mg/kg body weight during the dosing period, we assume 100% absorption and then use a set of standard values for each sex of each species, including body weight and average intake per day (Table 1 in the file “Documentation_for_Excel.pdf”). Using standard values, the daily dose-rate is calculated as follows:

$$\text{Dose rate} = (\text{dose} \times \text{intake/day} \times \text{number of doses/week}) / (\text{animal weight} \times 7 \text{ days/week})$$

or, equivalently,

$$\text{Dose rate} = \text{dose} \times \text{intake/day as a proportion of body weight} \times \text{proportion of week dose is administered}$$

In an experiment where the animals were dosed the entire time on test this value would equal the average daily dose rate.

In many experiments the administration of the test compound is stopped before the terminal sacrifice or before the death of the last animal. In such cases, averaging the dose over the course of the experiment will lower the daily dose level. By convention we take the total dose administered and spread this over the entire experimental period. If the pattern of dosing varied for a given dose group or the dose amount was changed, this is taken into account in the calculated dose. See the description of the dataset cpdb.supvary.xls in the Supplementary Database of Dosing and Survival documentation “CPDB_Supplementary.pdf”.

II. G. Carcinogenic Potency (TD₅₀)

A numerical description of carcinogenic potency, the TD₅₀ is estimated for each set of tumor incidence data reported in the CPDB, thus providing a standardized quantitative measure for comparisons and analyses of many issues in carcinogenesis. In a simplified way, TD₅₀ may be defined as follows: for a given target site(s), if there are no tumors in control animals, then TD₅₀ is that chronic dose-rate in mg/kg body wt/day which would induce tumors in half the test animals at the end of a standard lifespan for the species. Since the tumor(s) of interest often does occur in control animals, TD₅₀ is more precisely defined as: that dose-rate in mg/kg body wt/day which, if administered chronically for the standard lifespan of the species, will halve the probability of remaining tumorless throughout that period. TD₅₀ is analogous to LD₅₀, and a low value of TD₅₀ indicates a potent carcinogen, whereas a high value indicates a weak one. TD₅₀ can be computed for any particular type of neoplasm, for any particular tissue, or for any combination of these.

For NCI/NTP bioassays the TD₅₀, is confidence limits, shape of the dose-response and statistical significance are based on lifetable incidence data. For the general literature (cpdb.lit.xls) the TD₅₀, shape of the dose-response and statistical significance are based on summary incidence data. Wherever results were available in the published paper, the summary analysis uses the number of animals alive at the time of death of the first animal with the tumor of interest. If that is not available in the paper, then the number examined histologically is used. If only the number at the start of the experiment is available, then that is used.

II. H. Statistical Significance and Dose Response

Data on shape of the dose response and statistical significance are generated during the estimation of carcinogenic potency (TD₅₀) for the CPDB. These fields are included in the CPDB because they are relevant to the criteria for inclusion of tissue-tumor combinations reported in NCI/NTP Technical Reports. For the CPDB, tissue-tumor combinations that are reported in the statistical tables of the Technical Report but do not have an opinion by NCI/NTP about carcinogenic activity are in the CPDB if the slope of the dose response is different from zero ($p < 0.05$). For results from the general literature the significance of the dose response does not affect the inclusion of particular sites; inclusion is based on the χ^2 test comparing each dose group with controls. For completeness, curve and statistical significance are included in the literature dataset cpdb.lit.xls as in cpdb.ncintp.xls. See “Selection of Tissue and Tumor Types for Inclusion” (Section II.C, above).

The shape of the dose-response is based on the χ^2 goodness-of-fit statistic to test the validity of a linear relationship between dose and tumor incidence. If the goodness-of-fit test indicated statistically significant departure from linearity ($p < 0.05$) and this departure was downward, the analysis was repeated without the highest dose group. If there was an upward departure from linearity, no groups were eliminated. See the field “curve” in the file “Documentation_for_Excel.pdf” and the Appendix to the file CPDB.pdf.

For statistical significance, the likelihood ratio statistic tests the hypothesis that the test agent has no carcinogenic effect, i.e. the statistical significance associated with testing whether the slope of the dose-response is different from zero. If there was a significant downward curvature in the dose-response and the highest dose group was eliminated the statistical significance is based on the reduced number of dose groups. See the field “pval” in the file “Documentation_for_Excel.pdf” and the Appendix to the file CPDB.pdf.

III. Excel Linking Instructions

There are 14 Excel files. Codes used throughout the CPDB are defined in a set of appendix datasets listed below. Three additional appendices are given in text files: “Bibliography.pdf”, “Note_codes.pdf”, and “Monkey_documentation.pdf”.

The 14 datasets are:

Dataset Name	Brief Description
Main datasets	
cpdb.ncintp.xls	All data for NCI/NTP except doses and incidence
cpdb.ncntdose.xls	Dose and incidence data for NCI/NTP
cpdb.lit.xls	All data for literature except: 1) doses and incidence (cpdb.litdose.xls) 2) supplementary data reported in cpdb.supp.xls and cpdb.suppvary.xls
cpdb.litdose.xls	Dose and incidence data for literature
cpdb.chemname.xls	Chemical names, three-letter identification codes and CAS numbers
cpdb.cit.xls	Brief citation to published paper in literature
Datasets of supplementary literature data	
cpdb.supp.xls	Experimental protocol information for a dose group as reported in published paper
cpdb.suppvary.xls	Variable dosing information for a dose group
Datasets of code definitions	
cpdb.species.xls	Species code definitions
cpdb.route.xls	Route code definitions
cpdb.strain.xls	Strain code definitions
cpdb.tissue.xls	Tissue code definitions
cpdb.tumor.xls	Tumor histopathology code definitions
cpdb.journal.xls	Journal code definitions

A. Field Descriptions

A complete description of each field in the Excel files can be found in the file “Documentation_for_Excel.pdf”.

The rows in the CPDB represent different units of analysis depending upon the dataset. For the cpdb.ncintp.xls and cpdb.lit.xls datasets, each row represents a tissue-tumor combination within an experiment. For the cpdb.ncntdose.xls, cpdb.litdose.xls and cpdb.supp.xls datasets, each row represents a dose-group within an experiment. For the cpdb.suppvary.xls dataset, each row represents a dosing protocol within a dose level in an experiment. The cpdb.suppvary.xls dataset only contains data for variable dosing protocols. When there is variable dosing, fields in the cpdb.supp.xls dataset that correspond to fields in the cpdb.suppvary.xls dataset are blank: expdose, xpo, xpowk, hrpxpo, mlxpo, and age. For the cpdb.cit.xls dataset, each row represents a literature citation. For the remaining datasets, each row represents a code definition: cpdb.chemname.xls, cpdb.species.xls, cpdb.route.xls, cpdb.strain.xls, cpdb.tissue.xls, cpdb.tumor.xls, and cpdb.journal.xls.

B. Structure of the Excel Files

The datasets cpdb.ncintp.xls and cpdb.lit.xls contain all data except 1) doses and incidences, which are in the datasets cpdb.ncntdose.xls and cpdb.litdose.xls and 2) experimental protocol data contained in cpdb.supp.xls and cpdb.suppvary.xls. Each row in the files cpdb.ncintp.xls and cpdb.lit.xls represents a tissue-tumor combination. The datasets cpdb.ncintp.xls and cpdb.lit.xls contain information on each chemical, species, strain, sex, route, and published paper.

There is a one-to-many mapping between a row in cpdb.ncintp.xls and the rows in cpdb.ncntdose.xls. The “idnum” field is the key that maps the tissue-tumor combination to its associated doses. The same idea applies to cpdb.lit.xls and to cpdb.litdose.xls.

Linkages among the datasets utilize several fields described in the table below:

Each entry in the “chemcode” field in the cpdb.ncintp.xls and cpdb.lit.xls datasets allows one to link to the “chemcode” field in the cpdb.chemname.xls datasets, which also contains CAS numbers.

Each entry in the “papernum” field of the cpdb.lit.xls dataset allows one to link to the “papernum” field in the cpdb.cit.xls dataset, which contains the abbreviated citation information.

The cpdb.supp.xls dataset can be linked to the cpdb.lit.xls dataset using the fields in common between them that define an experiment: chemcode, species, sex, strain, route, and papernum.

The cpdb.supp.xls dataset can be linked to the cpdb.suppvary.xls dataset using the idvary field. The field “varynum” in cpdb.suppvary.xls indicates the order in which the doses were administered for each “idvary” number. Fewer than 20% of the experiments in cpdb.lit.xls have variable dosing, and are therefore included in cpdb.suppvary.xls.

Organization of dataset linking

Dataset	Dataset	Linking fields
cpdb.lit.xls	cpdb.litdose.xls	idnum
cpdb.lit.xls	cpdb.supp.xls	chemcode, species, sex, strain, route, papernum
cpdb.lit.xls	cpdb.cit.xls	papernum
cpdb.lit.xls	cpdb.species.xls	species
cpdb.lit.xls	cpdb.route.xls	route
cpdb.lit.xls	cpdb.strain.xls	strain
cpdb.lit.xls	cpdb.tissue.xls	tissue
cpdb.lit.xls	cpdb.tumor.xls	tumor
cpdb.ncintp.xls	cpdb.ncntdose.xls	idnum
cpdb.ncintp.xls	cpdb.species.xls	species
cpdb.ncintp.xls	cpdb.route.xls	route
cpdb.ncintp.xls	cpdb.strain.xls	strain
cpdb.ncintp.xls	cpdb.tissue.xls	tissue (subfield of 3 characters in cpdb.ncintp.xls)
cpdb.ncintp.xls	cpdb.tumor.xls	tumor (subfield of 3 characters in cpdb.ncintp.xls)
cpdb.cit.xls	cpdb.journal.xls	subfield of citation in cpdb.cit.xls; jcode in cpdb.journal.xls
cpdb.supp.xls	cpdb.suppvary.xls	idvary

Table 1. Standard Values For Dose Calculation: Animal Lifespans, Weights, And Intake By Diet, Water, And Inhalation^A

Experimental Animal	Sex	Standard lifespan (year)^b	Weight (kg)^c	Food (g)^c	Food as % body weight/day	Water (mL/day)^d	Inhalation volume (L/min)^e
Rodents							
Mouse	Male	2	0.03	3.6	12.00		0.03
	Female	2	0.025	3.25	13.00		0.03
Rat	Male	2	0.5	20	4.00		0.10
	Female	2	0.35	17.5	5.00		0.10
Hamster	Male	2	0.125	11.5	9.20		0.06
	Female	2	0.110	11.5	10.45		0.06
Monkeys							
Cynomolgus (<i>Macaca fascicularis</i>)	Both	20					
Rhesus (<i>Macaca mulatta</i>)	Both	20					
Prosimians							
Bush babies (<i>Galago crassicaudatus</i>)	Both	10					
Tree shrews (<i>Tupaia glis</i>)	Both	4.5					
Dog	Both	11	16	400			

^aAlthough values sometimes vary depending on the source, those given here are within reasonable limits of those usually found in the published literature. No value is given when this information was not necessary for our dose calculation.

^bRat and mouse: based on NCI trichloroethylene bioassay (NCI, 1976); hamster: data of Williams (1976); nonhuman primates: data of S. M. Sieber (Laboratory of Chemical Pharmacology, NCI, National Institute of Health, Bethesda, MD), personal communication; bush babies: ages adapted from Dittmer (1973); tree shrews: data of D. J. Reddy (Northwestern University, Chicago, IL), personal communication; dog: data of M. S. Redfearn (Division of Animal Resources, University of California, Berkeley), personal communication.

^cRat and mouse: based on NCI trichloroethylene bioassay (NCI, 1976); hamster and dog: data of D. Brooks (University of California, Davis), personal communication.

^dMouse, rat and dog: data from NIOSH (Sweet, 1993); hamster: data from Hoeltge, Inc.

^eMouse: data of Sanockij (1971); rat: data of Baker *et al.* (1979); hamster: data of Guyton (1947).