US EPA TOXCAST TOXREFDB DATA RELEASE OCTOBER 2014

This file describes the contents of the October 2014 ToxCast ToxRefDB data release. The zip file contains the following summary-level file, not including this README file:

[1] "toxrefdb_study_endpoint_summary_AUG2014_FOR_PUBLIC_RELEASE.csv"
[2] "toxrefdb_study_tg_effect_summary_AUG2014_FOR_PUBLIC_RELEASE.csv"
[3] "toxrefdb_endpoint_matrix_AUG2014_FOR_PUBLIC_RELEASE.csv"

FILE: toxrefdb_study_endpoint_summary_AUG2014_FOR_PUBLIC_RELEASE.csv

Description: This summary file is intended to provide No Effect Level (NEL)/Lowest Effect Level (LEL) and/or No Observed Adverse Effect Level (NOAEL)/Lowest Observed Adverse Effect Level (LOAEL) across the thousands of studies currently in ToxRefDB. This is the first time NEL and NOAEL values have been derived from the database and released. NEL and NOAEL are not explicitly captured in ToxRefDB but are calculated by definition as the next lower dose level from the LEL or LOAEL. For any single study, multiple N(OA)EL/L(OA)EL are captured based on the endpoint categories relevant to the study type. For example, a multigenerational reproductive study attempts to achieve a parental, offspring, and reproductive N(OA)EL/L(OA)EL. Therefore, any single row in this file will represent the N(OA)EL/L(OA)EL for a particular study-endpoint_category combination and should be interpreted as such.

chemical_id: DSSTox_GSID has direct match to ToxCast/Tox21 library
chemical_casrn: CAS Registry No.
chemical_name: Chemical name (generally should match with ToxCast chemical library files)
chemical_sets: Provides supplemental information in terms of which ToxCast testing phase
data_source: Primary source of study or study review (e.g., OPP DER = Office of Pesticide Program Data Evaluation Record, open_lit = open literature study)
entry_status_id: Field match with entry_status (key value used in database)
entry_status: Summarizes the status of data entry (partially complete (effect data) = data entry is complete but independent external review has not been completed; complete = indendent review has been completed (only a fraction of the studies have undergone this external review)
entry_level_id: Field match with entry_level (key value used in database)
entry_level: Level of detail the study was entered (for the vast majority of studies all treatment related effects were entered); summary = all effects entered provided in the summary document or abstract but no assurance that all effects were mentioned

usability: 1&2= Guideline Acceptable; 3=Non-Guideline Acceptable; 4=Unacceptable; 5=Incomplete/Deficient Report; 6=Not Evaluated (Recommendation
usability_desc: Description of data_usability_code
study_id: Unique study identifier (database generated)
source_study_numeric_id: Most often the MRID (EPA OPP study id) but can be PMID for Open Lit
source_study_alphanumeric_id: Other study id information or source (non-numeric) study id
year: Year study was conducted/concluded/report written
citation: Reference for study; only to be used as descriptive information and not definitive citation as these are not fully cross-referenced.
guideline_no: Closest match to OPPTS guideline number (Study type identifier)
guideline_name: Closest match to OPPTS guideline name (Study type identifier)
study_type_id: Field match with study_type (key value used in database)
species_id: Field match with species (key value used in database)
strain: Tested strain if exact match
comments_animal: usually this provides specification of strain
admin_method: Method of administration
admin_route: Route of administration
dose_start: Starting time (generally as listed in study report)
dose_start_unit: Unit of time including gestational days and generations
dose_end: Dose end time (generally as listed in study report)
dose_end_unit: Unit of time including gestational days and generations
lot_batch: Lot and batch information of test material if provided
purity: Purity of test material if provided
source: Source of test material if provided
ldt: low dose tested (in this file this is the same as min_dose)
hdt: high dose tested (in this file this is the same as max_dose)
dose_unit: Unit of dose
no_doses_tested: Number of dose groups (low-1, mid-2, mid-high-3, high-4 for example; this group male and female dosing to a single level)
study_type: Study type group based on guideline_no (CHR=chronic/cancer; MGR=multigenerational reproductive; DEV=Prenatal developmental; SUB=Subchronic; SAC=Subacute; REP=reproductive fertility)
species: Tested species
effect_category: Grouping of effects based on study type and endpoint type (generally every effect for a particular study type belongs to an endpoint category)
study_level_lel_dose_level: Used to determine if any LEL was established across any endpoint category for the study (this helps to

see if any effects were observed at all and to QC if effect data was entered at all)
lel_qualifier: Provides context for the LEL especially when no effect for a particular endpoint category was not observed.
lel_dose_level: Dose level of the LEL (Low-1,Mid-2,High-3…)
lel_dose: Dose of LEL
nel_qualifier: Provides context for the NEL; If the nel_dose is null (blank) then the user of this data can decide if they would like to derive the NEL based on the LEL/10 or LEL/3 or to leave blank/null/missing.
nel_dose_level: Dose level of the NEL
nel_dose: Dose of NEL (see nel qualifier for context around missing/null/blank values)
study_level_loael_dose_level: Used to determine if any LOAEL was established across any endpoint category for the study (this helps to see if any effects were observed at all and to QC if effect data was entered at all)
loael_qualifier: Provides context for the LOAEL especially when no effect for a particular endpoint category was not observed. NOAEL and LOAEL were derived primarily from the OPP source documents and most other data sources do not have NOAEL/LOAEL values.
loael_dose_level: Dose level of the LOAEL (Low-1,Mid-2,High-3…)
loael_dose: Dose of LOAEL
noael_qualifier: Provides context for the NOAEL; If the noael_dose is null (blank) then the user of this data can decide if they would like to derive the NOAEL based on the LOAEL/10 or LOAEL/3 or to leave blank/null/missing.
noael_dose_level: Dose level of the NOAEL
noael_dose: Dose of NOAEL (see nel qualifier for context around missing/null/blank values)

FILE: toxrefdb_study_tg_effect_summary_AUG2014_FOR_PUBLIC_RELEASE.csv

Description: ToxRefDB stores traditional in vivo toxicology study data in a relational database but this information can be collapsed with the understanding that certain information has been aggregated. The basic structure of ToxRefDB is that a chemical can have many studies, a study can have many treatment groups and that any treatment group can have many effects. This file contains a collapsed version of the chemical, study, treatment group and effect data. A particular row in this file with NULL values for a given treatment group means that no treatment related effects were observed. Otherwise each row in this file indicates a treatment related effect and tracks back to the treatment group and study information. This file is intended for more advanced users of the data or for subsetting all ToxRefDB data for a particular chemical. The fields of particular value for interpreting these data have been bolded.

chemical_id: DSSTox_GSID has direct match to ToxCast/Tox21 library
chemical_casrn: CAS Registry No.
chemical_name: Chemical name (generally should match with ToxCast chemical library files)
chemical_sets: Provides supplemental information in terms of which ToxCast testing phase
data_source: Primary source of study or study review (e.g., OPP DER = Office of Pesticide Program Data Evaluation Record, open_lit = open literature study)
entry_status_id: Field match with entry_status (key value used in database)
entry_status: Summarizes the status of data entry (partially complete (effect data) = data entry is complete but independent external review has not been completed; complete = indendent review has been completed (only a fraction of the studies have undergone this external review)
entry_level_id: Field match with entry_level (key value used in database)
entry_level: Level of detail the study was entered (for the vast majority of studies all treatment related effects were entered); summary = all effects entered provided in the summary document or abstract but no assurance that all effects were mentioned
usability: 1&2= Guideline Acceptable; 3=Non-Guideline Acceptable; 4=Unacceptable; 5=Incomplete/Deficient Report; 6=Not Evaluated (Recommendation
usability_desc: Description of data_usability_code
study_id: Unique study identifier (database generated)
source_study_numeric_id: Most often the MRID (EPA OPP study id) but can be PMID for Open Lit
source_study_alphanumeric_id: Other study id information or source (non-numeric) study id
year: Year study was conducted/concluded/report written
citation: Reference for study; only to be used as descriptive information and not definitive citation as these are not fully cross-referenced.
guideline_no: Closest match to OPPTS guideline number (Study type identifier)
guideline_name: Closest match to OPPTS guideline name (Study type identifier)
study_type_id: Field match with study_type (key value used in database)
study_type: Study type group based on guideline_no (CHR=chronic/cancer; MGR=multigenerational reproductive; DEV=Prenatal developmental; SUB=Subchronic; SAC=Subacute; REP=reproductive fertility)
species: Tested species
species_id: Field match with species (key value used in database)
strain: Tested strain if exact match
comments_animal: usually this provides specification of strain

admin_method: Method of administration
admin_route: Route of administration
dose_start: Starting time (generally as listed in study report)
dose_start_unit: Unit of time including gestational days and generations
dose_end: Dose end time (generally as listed in study report)
dose_end_unit: Unit of time including gestational days and generations
lot_batch: Lot and batch information of test material if provided
purity: Purity of test material if provided
source: Source of test material if provided
ldt: low dose tested (in this file this is the same as min_dose)
hdt: high dose tested (in this file this is the same as max_dose)
toxrefdb_study_dose_unit: Unit of dose
no_doses_tested: Number of dose groups (low-1, mid-2, mid-high-3, high-4 for example; this group male and female dosing to a single level)
tg_id: Treatment group id (unique database generated id)
generation: Generation of the treatment group
gender: Gender or gender group (Male and Female) of the Treatment group
dosing_period: Generally this is initial-to-terminal but can be an indicator of an interim sacrifice or other group
dose_level: Dose level of the particular treatment group (good for aggregating across gender as the actual reported dose will often be different due to varying food consumption rates
dose: The dose (primarily in mg/kg/day) consumed/inhaled by the treatment group (Minimum of this field for any given chemical-, study-, species-, effect- or endpoint-combination will allow for the derivation of a lowest effect level (LEL)
toxrefdb_tg_dose_unit: Unit of dose
duration: Duration in time of dosing (absolute time; does not include Gestational time for instance)
duration_unit: unit of dosing time; days, weeks, years
no_animals: Number of animals in treatment group
effect_id: Unique effect id (Database generated)
effect_type; effect_target; effect_desc: These three fields generally describe the observed effect. (See http://www.epa.gov/ncct/toxrefdb/ under publications for more detailed information)
effect_type_id; effect_target_id; effect_desc_id: Foreign key values (matched to effect_type, effect_target, effect_desc used in the database (potentially useful for indexing or merging))
direction: Direction of effect (generally increase or decrease)
effect_free_text: Free-text description of effect (often used to provide additional information if reported effect differs at all from standardized vocabulary)
target_site: Provides additional cell or region information
focal_diffuse: Focal or Diffuse (for pathology only)

loael: TRUE(-1) or FALSE (0) on whether or not the effect was part of a LOAEL criteria; If TRUE (primarily only captured for OPP-DER (EPA Pesticide Program Office Toxicity Reports)
effect_category: Grouping of effects based on study type and endpoint type (generally every effect for a particular study type belongs to an endpoint category)
endpoint_type: High level classification/grouping of effects
endpoint_system: Grouping of effects into general target organ systems
endpoint_target: Grouping of effect by target organ (similar to effect_target, but further groups in-life observations and clinical pathology effects)
endpoint_lifestage: lifestage in which effect was observed (adult, pregnant, juvenile, or fetal)

FILE: toxrefdb_endpoint_matrix_AUG2014_FOR_PUBLIC_RELEASE.csv

Description: The endpoint matrix file contains the chemical-endpoint-specific Lowest Effect Levels (LEL; mg/kg/day). The 1 Million values indicate that the endpoint was tested (or assumed to have been tested) and was not observed (or treatment-related) in the study. The NA values indicate that the chemical was not tested for that endpoint. The LEL is the lowest dose at which any effect in the endpoint was observed. For example, *CHR_rat_SystemicCarcinogenic* LEL essentially covers any systemic or carcinogenic effect observed in the chronic/cancer rat study. Whereas, *CHR_rat_SystemicCarcinogenic_adult_PathologyNeoplastic_AccessoryDigestive_Liver* describes the LEL for any liver neoplastic lesion, including adenomas and carcinomas. The endpoint matrix may be useful for modeling efforts and is not recommended for casual browsing of chemical-endpoint-specific effects. The 'X' column provides a concatenated set of chemical id and name information. This allows a user to load in the matrix file directly into a programming environment. The endpoints (columns) are concatenated combinations of the effect category and endpoint groupings as defined below:

effect_category: Grouping of effects based on study type and endpoint type (generally every effect for a particular study type belongs to an endpoint category)
endpoint_type: High level classification/grouping of effects
endpoint_system: Grouping of effects into general target organ systems
endpoint_target: Grouping of effect by target organ (similar to effect_target, but further groups in-life observations and clinical pathology effects)
endpoint_lifestage: lifestage in which effect was observed (adult, pregnant, juvenile, or fetal)