Creating a Database of Amphibian Dermal Exposure Data

Purucker ST, Snyder MN, Glinski DA, Van Meter RJ, Garber K, Cyterski MJ, Sinnathamby S, Henders

---

Introduction

The purpose of this script is to combine the data sets from Van Meter et al. (2014, 2015, 2016, 2018), Glinski et al. (2018a, b, 2019, 2020), and Henson-Ramsey (2008) to create a collated database of amphibian dermal exposure data.

Manuscript	Data Set (Original Source Link)	Data Set (Repo Link)	Additional Data Sets
Van Meter et al. 2014	good_data.csv	vm2014_data.csv	
Van Meter et al. 2015	good_data.csv	vm2014_data.csv	
Van Meter et al. 2016	RDATA.csv	vm2016_data.csv	
Van Meter et al. 2018	vm2017_merge.csv	vm2017_merge.csv	
Glinski et al. 2018a (dehydration)	dehydration3.csv	dag2016_data_dehydration.csv	
Glinski et al. 2018b (metabolites)	exposure_experiment.csv	dag2016_data_metabolites_4merge.csv	
Glinski et al. 2019 (biomarkers)	exposure_mixtures3.csv	dag2018_data_biomarkers.csv	biomarker.csv (dag_biomarker2.csv)
Glinski et al. 2020 (dermal routes)	Water_soil.csv	dag2019_dermal_routes.csv	Dermal_routes_weights.csv (weights)
Henson-Ramsey 2008	HensonRamseyetal2008_data.csv	hr2008_data.csv	

Computational environment

This repository can be found at: https://github.com/puruckertom/amphib_dermal_collation

If you are on a Mac and get xquartz complaints (knitr), install from: https://www.xquartz.org/

Data from Relevant Studies

Van Meter et al. 2014 and Van Meter et al. 2015

Van Meter et al. 2014 performed exposures for 5 pesticide active ingredients (imidacloprid, pendimethalin, atrazine, fipronil, tridimefon) and 7 species (Southern leopard frog (*Lithobates sphenocephala*), Fowler's toad (*Anaxyrus fowleri*), gray treefrog (*Hyla versicolor*), Northern cricket frog (*Acris crepitans*), Eastern narrowmouth toad (*Gastrophryne carolinensis*), barking treefrog (*Hyla gratiosa*) and green treefrog (*Hyla cinerea*)). Whole body tissue concentrations were measured after an 8 hour exposure period to contaminated soil. Pesticides were applied at the maximum legally allowable application rates scaled down to the area of a 10-gallon aquarium (1225 cm^2).

Van Meter et al. 2015 contrasted two pesticide exposure scenarios: direct exposure through aerial overspray and indirect exposure through soil. These scenarios tested the same 5 pesticide active ingredients and two of the species (barking treefrog (*Hyla gratiosa*) and green treefrog (*Hyla cinerea*)). Pesticides were applied at the maxium legally allowable application rates scaled down to the size of a 10-gallon aquarium, with the exception of pedimethalin which was applied at 30% of the permitted application rate. This was due to pedimethalin's insolubility in the limited solvent and the water volumes used in this study.

For our purposes, the Van Meter et al. 2015 essentially adds the aerial overspray exposures to the Van Meter et al. 2014 data set.

Note: this file does include metabolites into the total for the parents

Data Set Dimensions, Column Names, and Summary:

```
## [1] 474  23
```

```
##  [1] "Species"        "Sample"         "Chemical"       "Instrument"
##  [5] "good"           "Application"    "app_rate_g_cm2" "TissueConc"
##  [9] "SoilConc"       "logKow"         "BCF"            "bodyweight"
## [13] "initialweight"  "Solat20C_mgL"   "Solat20C_gL"    "molmass_gmol"
## [17] "Density_gcm3"   "AppFactor"      "SA_cm2"         "VapPrs_mPa"
## [21] "Koc_gmL"        "HalfLife_day"   "HabFac"
```

```
##    Species             Sample            Chemical          Instrument
##  Length:474         Length:474         Length:474         Length:474
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##       good   Application        app_rate_g_cm2    TissueConc
##  Min.   :1   Length:474         Min.   :0e+00   Min.   : 0.007484
##  1st Qu.:1   Class :character   1st Qu.:0e+00   1st Qu.: 0.246753
##  Median :1   Mode  :character   Median :0e+00   Median : 0.575811
##  Mean   :1                      Mean   :1e-05   Mean   : 1.908242
##  3rd Qu.:1                      3rd Qu.:2e-05   3rd Qu.: 1.743142
##  Max.   :1                      Max.   :2e-05   Max.   :23.441298
##                                 NA's   :151
##     SoilConc            logKow           BCF             bodyweight
##  Min.   : 0.00625   Min.   :0.570   Min.   :  0.0018   Min.   :0.5004
##  1st Qu.: 0.20866   1st Qu.:2.500   1st Qu.:  0.0755   1st Qu.:1.3162
##  Median : 3.49248   Median :3.110   Median :  0.2069   Median :1.8550
```

```
##  Mean   : 7.22468   Mean   :3.142   Mean   : 11.3804   Mean   :1.8658
##  3rd Qu.:10.06719   3rd Qu.:4.000   3rd Qu.:  1.0828   3rd Qu.:2.3489
##  Max.   :81.71115   Max.   :5.180   Max.   :396.8461   Max.   :3.9931
##
##  initialweight     Solat20C_mgL     Solat20C_gL       molmass_gmol
##  Min.   :0.5004   Min.   :  0.30   Min.   :0.00030   Min.   :215.7
##  1st Qu.:1.6614   1st Qu.:  3.78   1st Qu.:0.00378   1st Qu.:215.7
##  Median :2.1766   Median : 30.00   Median :0.03000   Median :291.7
##  Mean   :2.2307   Mean   :123.20   Mean   :0.12320   Mean   :299.5
##  3rd Qu.:2.7601   3rd Qu.:260.00   3rd Qu.:0.26000   3rd Qu.:291.7
##  Max.   :5.5480   Max.   :510.00   Max.   :0.51000   Max.   :437.1
##
##   Density_gcm3     AppFactor           SA_cm2          VapPrs_mPa
##  Min.   :1.170   Min.   :    850   Min.   : 0.7915   Min.   :0.00020
##  1st Qu.:1.187   1st Qu.:  47011   1st Qu.: 1.5393   1st Qu.:0.00037
##  Median :1.220   Median : 143055   Median : 1.7866   Median :0.02000
##  Mean   :1.288   Mean   : 291904   Mean   : 3.0232   Mean   :0.34774
##  3rd Qu.:1.480   3rd Qu.: 348598   3rd Qu.: 2.0882   3rd Qu.:0.04000
##  Max.   :1.543   Max.   :4490329   Max.   :23.3326   Max.   :4.00000
##                  NA's   :151
##     Koc_gmL        HalfLife_day       HabFac
##  Min.   :   122   Min.   : 26.00   Length:474
##  1st Qu.:   122   1st Qu.: 26.00   Class :character
##  Median :   520   Median : 80.00   Mode  :character
##  Mean   : 20406   Mean   : 70.85
##  3rd Qu.:   825   3rd Qu.: 84.00
##  Max.   :243000   Max.   :125.00
##
```

Van Meter et al. 2016

Van Meter et al. 2016 considered bioconcentration of 5 current-use pesticides (imidacloprid, atrazine, triadimefon, fipronil, and pedimethalin) in American toads (*Bufo americanus*) across soil types. Toads were exposed to one of two soil types with significantly different organic matter content (14.1% = high organic matter, 3.1% = low organic matter). Whole body tissue concentrations were measured after an 8 hour exposure period to contaminated soil. Pesticides were applied at the maximum legally allowable application rates scaled down to the area of six 0.94 L Pyrex glass bowls each with a 15 cm diameter.

Note: this file does include metabolites into the total for the parents

Data Set Dimensions, Column Names, and Summary:

```
## [1] 264  11
```

```
##  [1] "Day"         "Row"         "Column"      "Pesticide"   "SoilType"
##  [6] "BodyBurden"  "Soil"        "Weight"      "Total"       "Formulation"
## [11] "Parent"
```

```
##       Day             Row           Column           Pesticide
##  Min.   :0.000   Min.   :1.000   Length:264         Length:264
##  1st Qu.:2.000   1st Qu.:2.000   Class :character   Class :character
##  Median :2.000   Median :4.000   Mode  :character   Mode  :character
##  Mean   :2.326   Mean   :4.023
##  3rd Qu.:3.000   3rd Qu.:6.000
```

```
##  Max.   :3.000   Max.   :7.000
##    SoilType           BodyBurden           Soil              Weight
##  Length:264         Min.   :-0.0378   Min.   :-0.10518   Min.   : 6.964
##  Class :character   1st Qu.: 0.0486   1st Qu.: 0.02086   1st Qu.:10.524
##  Mode  :character   Median : 0.1099   Median : 1.49572   Median :11.740
##                     Mean   : 0.4955   Mean   : 6.02720   Mean   :12.044
##                     3rd Qu.: 0.3650   3rd Qu.: 8.64289   3rd Qu.:13.440
##                     Max.   : 6.8744   Max.   :39.57404   Max.   :23.340
##      Total         Formulation         Parent
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :1.0000
##  Mean   :0.3636   Mean   :0.4091   Mean   :0.5909
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
```

Glinski et al. 2018a (Dehydration)

Glinski et al. 2018a studied how amphibian hydration status influences uptake of pesticides through dermal exposure. Amphibians (Southern leopard frogs (*Lithobates sphenocephala*) and Fowler's toads (*Anaxyrus fowleri*)) were dehyrated for periods of 0, 2, 4, 6, 8, or 10 hours prior to exposure to pesticide-contaminated soils. Pesticides studied included atrazine, triadimefon, metolachlor, chlorothalonil, and imidacloprid. Soil and whole-body homogenates were measured after an 8 hour exposure period. Pesticides were applied at the maximum legally allowable application rates scaled down to the area of six 0.94 L Pyrex glass bowls each with a 15 cm diameter.

Note: this file does not combine daughters with parents
Note: this file has body burdens and soil concentrations as separate rows

Data Set Dimensions, column Names, and Summary:

```
## [1] 1494    8
```

```
## [1] "time"    "parent"  "analyte" "matrix"  "species" "conc"    "ID"
## [8] "weight"
```

```
##       time       parent            analyte            matrix
##  Min.   : 0   Length:1494        Length:1494        Length:1494
##  1st Qu.: 2   Class :character   Class :character   Class :character
##  Median : 5   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 5
##  3rd Qu.: 8
##  Max.   :10
##    species              conc               ID                weight
##  Length:1494        Min.   :  0.00000   Length:1494        Min.   :0.6821
##  Class :character   1st Qu.:  0.02215   Class :character   1st Qu.:1.6108
##  Mode  :character   Median :  0.08482   Mode  :character   Median :3.0890
##                     Mean   :  6.17646                      Mean   :3.0810
##                     3rd Qu.:  2.60007                      3rd Qu.:4.3124
##                     Max.   :238.15019                      Max.   :7.2481
```

Henson-Ramsey 2008

Henson-Ramsey 2008 tested the biological impact of exposure to malathion for tiger salamanders (*Ambystoma tigrinum*). Tiger salamanders were exposed to contaminated soils with 50 ug/cm^2 or 100 ug/cm^2 malathion and through ingestion of an earthworm exposed to contaminated soils with 200 ug/cm^2 malathion. For each exposure, the malathion application rate was sprayed onto the approximately 1200g of soil in the 1060cm^2 polyethylene cages. Tissue concentrations were assessed for five treatment groups: unexposed, exposed to 50 ug/cm^2 contaminated soil for 1 day, exposed to 50 ug/cm^2 for 2 days, exposed to 50 ug/cm^2 contaminated soil for 2 days and fed a contaminated worm on the first exposure day, and exposed to 100 ug/cm^2 contaminated soil for 2 days and fed a contaminated worm on the first exposure day.

Data Set Dimensions, Column Names, and Summary:

```
## [1]  9 12
```

```
##  [1] "chemical"        "species"         "tissue_conc_ugg" "sample_id"
##  [5] "body_weight_g"   "formulation"     "soil_type"       "application"
##  [9] "app_rate_g_cm2"  "exp_duration"    "soil_conc_ugg"   "source"
```

```
##    chemical           species          tissue_conc_ugg  sample_id
##  Length:9           Length:9           Min.   :0.050   Length:9
##  Class :character   Class :character   1st Qu.:0.350   Class :character
##  Mode  :character   Mode  :character   Median :1.420   Mode  :character
##                                        Mean   :1.186
##                                        3rd Qu.:1.470
##                                        Max.   :3.730
##  body_weight_g   formulation    soil_type      application
##  Min.   :20.89   Mode:logical   Mode:logical   Length:9
##  1st Qu.:44.15   NA's:9         NA's:9         Class :character
##  Median :46.26                                 Mode  :character
##  Mean   :43.73
##  3rd Qu.:48.93
##  Max.   :50.92
##  app_rate_g_cm2   exp_duration soil_conc_ugg     source
##  Min.   :5e-05   Min.   :24    Mode:logical   Length:9
##  1st Qu.:5e-05   1st Qu.:24    NA's:9         Class :character
##  Median :5e-05   Median :48                   Mode  :character
##  Mean   :5e-05   Mean   :40
##  3rd Qu.:5e-05   3rd Qu.:48
##  Max.   :5e-05   Max.   :48
```

Glinski et al. 2018b (Metabolites)

Glinski et al. 2018b assessed the potential metabolic activation of pesticides (atrazine, triadimefon, fopronil) in amphibians. This data set (1) contains *in vitro* and *in vivo* metabolic rate constants derived from toad (*Anaxyrus terrestris*) livers during experiments measuring the depletion of pesticides and the formation of their metabolites. Pesticides were applied at the maximum legally allowable application rates scaled down to the area of a 10-gallon aquarium (1225 cm^2).

Metabolites Data Set (1) Data Set Dimensions, Column Names, and Summary:

```
## [1] 352   6
```

```
## [1] "time"      "parent"    "analyte"   "matrix"    "conc"      "replicate"
```

```
##       time          parent            analyte            matrix
##  Min.   : 0.00   Length:352         Length:352         Length:352
##  1st Qu.: 2.00   Class :character   Class :character   Class :character
##  Median :12.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   :16.41
##  3rd Qu.:24.00
##  Max.   :48.00
##       conc            replicate
##  Min.   :-0.01244   Min.   :1.00
##  1st Qu.: 0.01292   1st Qu.:1.75
##  Median : 0.08373   Median :2.50
##  Mean   : 2.12963   Mean   :2.50
##  3rd Qu.: 0.97824   3rd Qu.:3.25
##  Max.   :32.47385   Max.   :4.00
```

The *in vitro* derived constants were assessed for their precitability by exposing Fowler's toads (*Anaxyrus fowleri*) to contaminated soils at maximum application rate for 2, 4, 12, and 48 hours. This data set (merged) contains the data from the Fowler's toad experiment along with the tissue concentrations from data set 1; this data set (merged) is used in subsequent steps.

Metabolites Data Set (merged) Data Set Dimensions, Column Names, and Summary:

```
## [1] 60 12
```

```
##  [1] "exp_duration"    "chemical"        "tissue_conc_ugg" "sample_id"
##  [5] "soil_type"       "app_rate_g_cm2"  "soil_conc_ugg"   "body_weight_g"
##  [9] "formulation"     "species"         "application"     "source"
```

```
##   exp_duration   chemical         tissue_conc_ugg    sample_id
##  Min.   : 2    Length:60          Min.   :0.08328   Length:60
##  1st Qu.: 4    Class :character   1st Qu.:0.33733   Class :character
##  Median :12    Mode  :character   Median :0.86010   Mode  :character
##  Mean   :18                       Mean   :1.42634
##  3rd Qu.:24                       3rd Qu.:1.88383
##  Max.   :48                       Max.   :7.62649
##  soil_type      app_rate_g_cm2      soil_conc_ugg  body_weight_g
##  Mode:logical   Min.   :1.100e-06   Mode:logical   Min.   :0.1879
##  NA's:60        1st Qu.:1.100e-06   NA's:60        1st Qu.:0.5925
##                 Median :2.700e-06                  Median :0.7144
##                 Mean   :9.237e-06                  Mean   :0.7350
##                 3rd Qu.:2.290e-05                  3rd Qu.:0.8782
##                 Max.   :2.290e-05                  Max.   :1.4909
##   formulation   species          application          source
##  Min.   :0    Length:60          Length:60          Length:60
##  1st Qu.:0    Class :character   Class :character   Class :character
##  Median :0    Mode  :character   Mode  :character   Mode  :character
##  Mean   :0
##  3rd Qu.:0
##  Max.   :0
```

Glinski et al. 2019 (Biomarkers)

Glinski et al. 2019 exposed Southern leopard frogs (*Lithobates sphenocephala*) to either the maximum or 1/10th maximum pesticide application rate to single, double, or triple pesticide mixtures of bifenthrin, metolachlor, and triadimefon to consider the typical co-application of pesticides during agricultural growing seasons. Tissue concentrations and metabolomic profiling of amphibian livers were studied after an 8 hour exposure period to pesticide-contaminated soil. Pesticides application rates were scaled down to the area of eight 0.94 L Pyrex glass bowls each with a 15 cm diameter.

Data Set Dimensions, Column Names, and Summary:

```
## [1] 192   9
```

```
## [1] "group"       "met"         "tdt"         "bif"         "frog.weight"
## [6] "sample_id"   "pesticide"   "rate"        "conc"
```

```
##     group                met                tdt                bif
##  Length:192         Min.   :-1.0000   Min.   :-1.0000   Min.   :-1.0000
##  Class :character   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.0000
##  Mode  :character   Median : 1.0000   Median : 1.0000   Median : 1.0000
##                     Mean   : 0.3333   Mean   : 0.3333   Mean   : 0.3333
##                     3rd Qu.: 1.0000   3rd Qu.: 1.0000   3rd Qu.: 1.0000
##                     Max.   : 1.0000   Max.   : 1.0000   Max.   : 1.0000
##   frog.weight     sample_id          pesticide            rate
##  Min.   :1.012   Length:192         Length:192         Length:192
##  1st Qu.:2.745   Class :character   Class :character   Class :character
##  Median :3.142   Mode  :character   Mode  :character   Mode  :character
##  Mean   :3.299
##  3rd Qu.:3.789
##  Max.   :6.739
##       conc
##  Min.   : 0.001061
##  1st Qu.: 0.069055
##  Median : 0.212920
##  Mean   : 0.801643
##  3rd Qu.: 0.521471
##  Max.   :19.879783
```

Van Meter et al. 2018 (Multiple Pesticides Study)

Van Meter et al. 2018 evaluated risks to amphibians after exposure to a single pesticide and pesticide mixtures. The five pesticides studied were three herbicides (atrazine, metolachlor, and 2,4-D), one insecticide (malathion), and one fungicide (propiconazole). Juvenile green frogs (*Lithobates clamitans*) were exposed to contaminated soils for 8 hours and metabolic analysis of amphibian livers was conducted to measure the effects. Pesticides were applied at the maximum legally allowable application rates individually and in mixtures of two or three pesticides within an herbicide or mixed pesticide group, scaled down to the area of six 0.94 L Pyrex glass bowls each with a 15 cm diameter.

Two data sets were generated from this study, one containing data for exposure to herbicides (single and mixed) and the other containing data for exposure to mixed pesticide treatments (herbicides, insecticide, fungicide).

Herbicide Data Set Data Set Dimensions, Column Names, and Summary:

```
## [1] 378  10
```

```
##  [1] "Group"     "ATZ"       "D"         "ME"        "AppRate"   "Weight"
##  [7] "SA"        "Media"     "Pesticide" "Conc"
```

```
##     Group                ATZ                D                 ME
##  Length:378         Min.   :-1.0000   Min.   :-1.0000   Min.   :-1.0000
##  Class :character   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.0000
##  Mode  :character   Median : 1.0000   Median : 1.0000   Median : 1.0000
##                     Mean   : 0.1429   Mean   : 0.1429   Mean   : 0.1429
##                     3rd Qu.: 1.0000   3rd Qu.: 1.0000   3rd Qu.: 1.0000
##                     Max.   : 1.0000   Max.   : 1.0000   Max.   : 1.0000
##     AppRate          Weight             SA           Media
##  Min.   :14.30   Min.   :0.9634   Min.   :1.107   Length:378
##  1st Qu.:23.60   1st Qu.:1.6929   1st Qu.:1.534   Class :character
##  Median :37.90   Median :2.0637   Median :1.720   Mode  :character
##  Mean   :39.31   Mean   :2.0892   Mean   :1.715
##  3rd Qu.:54.50   3rd Qu.:2.4927   3rd Qu.:1.919
##  Max.   :68.80   Max.   :3.6843   Max.   :2.406
##   Pesticide              Conc
##  Length:378         Min.   : 0.00000
##  Class :character   1st Qu.: 0.00000
##  Mode  :character   Median : 0.06358
##                     Mean   : 5.64721
##                     3rd Qu.: 1.46036
##                     Max.   :76.03573
```

Mixed Pesticide Data Set Data Set Dimensions, Column Names, and Summary:

```
## [1] 216   9
```

```
## [1] "Group"     "ATZ"       "MA"        "PROP"      "Pesticide" "Media"
## [7] "Conc"      "Weight"    "SA"
```

```
##     Group                ATZ                MA                PROP
##  Length:216         Min.   :-1.0000   Min.   :-1.0000   Min.   :-1.0000
##  Class :character   1st Qu.:-1.0000   1st Qu.:-1.0000   1st Qu.:-1.0000
##  Mode  :character   Median : 1.0000   Median : 1.0000   Median : 1.0000
##                     Mean   : 0.3333   Mean   : 0.3333   Mean   : 0.3333
##                     3rd Qu.: 1.0000   3rd Qu.: 1.0000   3rd Qu.: 1.0000
##                     Max.   : 1.0000   Max.   : 1.0000   Max.   : 1.0000
##   Pesticide            Media                Conc              Weight
##  Length:216         Length:216         Min.   : 0.00024   Min.   :1.188
##  Class :character   Class :character   1st Qu.: 0.32682   1st Qu.:1.786
##  Mode  :character   Mode  :character   Median : 1.61181   Median :2.014
##                                        Mean   : 5.46049   Mean   :2.203
##                                        3rd Qu.: 9.99874   3rd Qu.:2.455
##                                        Max.   :71.52122   Max.   :4.014
##        SA
##  Min.   :1.447
```

```
##  1st Qu.:1.833
##  Median :1.965
##  Mean   :2.047
##  3rd Qu.:2.203
##  Max.   :2.929
```

The herbicide and mixed pesticide data sets were cleaned prior and joined into a merged data set (referred to as Van Meter et al. 2018 Multiple Pesticides Study in subsequent steps). The single and mixed-pesticide treatments that were retained in the merged data set include atrazine, propiconazole, 2,4-D, malathion, and metolachlor. Original columns from the herbicide and mixed pesticide data sets were altered for standardization. These standardized columns will be used in future data cleaning steps in order to merge all data sets.

Merged Data Set Data Set Dimensions, Column Names, and Summary:

```
## [1] 137  12
```

```
##  [1] "app_rate_g_cm2"  "body_weight_g"   "chemical"        "tissue_conc_ugg"
##  [5] "sample_id"       "source"          "application"     "exp_duration"
##  [9] "formulation"     "soil_conc_ugg"   "soil_type"       "species"
```

```
##  app_rate_g_cm2      body_weight_g      chemical         tissue_conc_ugg
##  Min.   :2.600e-06   Min.   :0.9634   Length:137         Min.   : 0.00054
##  1st Qu.:1.430e-05   1st Qu.:1.7623   Class :character   1st Qu.: 0.27576
##  Median :2.360e-05   Median :2.0136   Mode  :character   Median : 1.41009
##  Mean   :2.004e-05   Mean   :2.1086                      Mean   : 7.36154
##  3rd Qu.:2.590e-05   3rd Qu.:2.3395                      3rd Qu.: 9.95084
##  Max.   :3.090e-05   Max.   :4.0141                      Max.   :72.62672
##   sample_id            source          application         exp_duration
##  Length:137         Length:137         Length:137         Min.   :8
##  Class :character   Class :character   Class :character   1st Qu.:8
##  Mode  :character   Mode  :character   Mode  :character   Median :8
##                                                           Mean   :8
##                                                           3rd Qu.:8
##                                                           Max.   :8
##   formulation soil_conc_ugg  soil_type        species
##  Min.   :0    Mode:logical   Mode:logical   Length:137
##  1st Qu.:0    NA's:137       NA's:137       Class :character
##  Median :0                                  Mode  :character
##  Mean   :0
##  3rd Qu.:0
##  Max.   :0
```

Glinski et al. 2020 (Dermal Routes)

Glinski et al. 2020 assessed dermal uptake in amphibians from exposure to three pesticides (bifenthrin, chlorpyrifos, trifloxystrobin). Pesiticide body burdens and hepatic metabolome for Leopard frogs were measured for two routes of uptake: uptake from contaminated soils versus uptake from contaminated surface water. Pesticides were applied at 1 ppm, scaled down to the area of eight 0.94 L Pyrex glass bowls each with a 15 cm diameter.

Data Set Dimensions, Column Names, and Summary:

```
## [1] 192   5
```

```
## [1] "Sample.ID"     "Analyte"       "Media"         "Matrix"
## [5] "Concentration"
```

```
##   Sample.ID           Analyte             Media              Matrix
##  Length:192         Length:192         Length:192         Length:192
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Concentration
##  Min.   :0.00000
##  1st Qu.:0.01036
##  Median :0.15326
##  Mean   :0.33962
##  3rd Qu.:0.44162
##  Max.   :3.40759
```

Application Rates

The table below concisely displays the pesticide applications rates (ug/cm^2) used in each relevant study as well as the variables used to compute the application rates.

Cleaning and Merging the Data Sets

Each data set was cleaned for merging. This consisted of dropping unneeded columns and standardizing column names of retained columns. Four columns were added to all data sets (soil type, formulation, exposure duration, and research study source).
Once each data set was cleaned, a local copy was saved and the data set was merged with the previously cleaned data sets.

The process of cleaning and merging each data set is briefly described below.

Van Meter et al. 2014/2015

Metabolites and parents that do not include metabolites were dropped from the data set. This includes atrazine, deisopropyl atrazine, desethyl atrazine, fipronil, fipronil-sulfone, triadimefon, triadimenol.

```
# drop metabolites and parents that do not include metabolites
vm2015_chem_drop <- c("Atrazine","Deisopropyl Atrazine","Desethyl Atrazine","Fipronil","Fipronil-Sulfon
chem_vector_drop <- which(vm2015$Chemical %in% vm2015_chem_drop)
vm2015_subset1 <- vm2015[-chem_vector_drop,]
vm2015_subset2 <- droplevels(vm2015_subset1)
```

pesticide	app_rate_ug_cm2	applied_mL	container	area_cm2	tota
Van Meter et al. 2014/2015					
atrazine	22.9000	75 MeOH	10-gal aquarium	1225	
fipronil	1.1000	75 MeOH	10-gal aquarium	1225	
imidacloprid	5.7000	75 MeOH	10-gal aquarium	1225	
pendimethalin	19.8000	75 MeOH	10-gal aquarium	1225	
triadimefon	2.7000	75 MeOH	10-gal aquarium	1225	
Van Meter et al. 2016					
atrazine	22.9000	75 MeOH	.94 L bowl	225*6	
fipronil	1.1000	75 MeOH	.94 L bowl	225*6	
imidacloprid	5.7000	75 MeOH	.94 L bowl	225*6	
pendimethalin	69.8000	75 MeOH	.94 L bowl	225*6	
triadimefon	2.7000	75 MeOH	.94 L bowl	225*6	
Van Meter et al. 2018					
atrazine	23.6000	50 MeOH	.94 L bowl	225*6	
2,4-D	14.3000	50 MeOH	.94 L bowl	225*6	
metolachlor	30.9000	50 MeOH	.94 L bowl	225*6	
malathion	25.9000	50 MeOH	.94 L bowl	225*6	
propiconazole	2.6000	50 MeOH	.94 L bowl	225*6	
Henson-Ramsey et al. 2008					
malathion	50.0000	NA	cage	1060	
Glinski et al. 2018a					
atrazine	23.9500	75 MeOH	.94 L bowl	225*6	
chlorothalonil	44.3000	75 MeOH	.94 L bowl	225*6	
imidacloprid	5.3900	75 MeOH	.94 L bowl	225*6	
metolachlor	31.0100	75 MeOH	.94 L bowl	225*6	
triadimefon	2.9100	75 MeOH	.94 L bowl	225*6	
Glinski et al. 2018b					
atrazine	22.9000	75 MeOH	10-gal aquarium	1225	
fipronil	1.1000	75 MeOH	10-gal aquarium	1225	
triadimefon	2.7000	75 MeOH	10-gal aquarium	1225	
Glinski et al. 2019					
bifenthrin (max)	3.4500	75 MeOH	.94 L bowl	225*8	
metolachlor (max)	30.6200	75 MeOH	.94 L bowl	225*8	
triadimefon (max)	2.8700	75 MeOH	.94 L bowl	225*8	
bifenthrin (1/10 max)	0.3450	75 MeOH	.94 L bowl	225*8	
metolachlor (1/10 max)	3.0620	75 MeOH	.94 L bowl	225*8	
triadimefon (1/10 max)	0.2870	75 MeOH	.94 L bowl	225*8	
Glinski et al. 2020					
bifenthrin	0.2889	400ml of 1ppm pesticide in water	.94 L bowl	225*8	
chlorpyrifos	0.3111	400ml of 1ppm pesticide in water	.94 L bowl	225*8	
trifloxystrobin	0.3022	400ml of 1ppm pesticide in water	.94 L bowl	225*8	

There were 278 observations with these chemicals. After dropping the 278 observations from the initial 474, the updated dimensions are:

```
## [1] 196  23
```

There were 15 unneeded columns dropped and 4 added for standarization.

```
# drop unneeded columns for merging
all_cols <- colnames(vm2015_subset2)
drop_cols <- c("Instrument", "good", "logKow", "BCF", "initialweight",
            "Solat20C_mgL", "Solat20C_gL", "molmass_gmol", "Density_gcm3","AppFactor", "SA_cm2", "VapPr
            "Koc_gmL", "HalfLife_day", "HabFac")
vm2015_subset3 <- vm2015_subset2[,!(names(vm2015_subset2) %in% drop_cols)]
colnames(vm2015_subset3)
```

```
## [1] "Species"        "Sample"         "Chemical"       "Application"
## [5] "app_rate_g_cm2" "TissueConc"     "SoilConc"       "bodyweight"
```

```
# add columns
soil_type <- c(rep("PLE",nrow(vm2015_subset3)))
formulation <- (rep(0,nrow(vm2015_subset3)))
exp_duration<- (rep(8,nrow(vm2015_subset3)))
source <- c(rep("rvm2015",nrow(vm2015_subset3)))
vm2015_subset4 <- cbind(vm2015_subset3, formulation, soil_type, exp_duration, source)
# standardize column names
colnames(vm2015_subset4)
```

```
##  [1] "Species"        "Sample"         "Chemical"       "Application"
##  [5] "app_rate_g_cm2" "TissueConc"     "SoilConc"       "bodyweight"
##  [9] "formulation"    "soil_type"      "exp_duration"   "source"
```

```
colnames(vm2015_subset4)[which(colnames(vm2015_subset4)=="Sample")]<-"sample_id"
colnames(vm2015_subset4)[which(colnames(vm2015_subset4)=="Species")]<-"species"
colnames(vm2015_subset4)[which(colnames(vm2015_subset4)=="Chemical")]<-"chemical"
colnames(vm2015_subset4)[which(colnames(vm2015_subset4)=="Application")]<-"application"
colnames(vm2015_subset4)[which(colnames(vm2015_subset4)=="TissueConc")]<-"tissue_conc_ugg"
colnames(vm2015_subset4)[which(colnames(vm2015_subset4)=="SoilConc")]<-"soil_conc_ugg"
colnames(vm2015_subset4)[which(colnames(vm2015_subset4)=="bodyweight")]<-"body_weight_g"
colnames(vm2015_subset4)
```

```
##  [1] "species"         "sample_id"       "chemical"        "application"
##  [5] "app_rate_g_cm2"  "tissue_conc_ugg" "soil_conc_ugg"   "body_weight_g"
##  [9] "formulation"     "soil_type"       "exp_duration"    "source"
```

```
# reorder vm2015 alphabetically
vm2015_merge <- vm2015_subset4[,order(names(vm2015_subset4))]

# write a local copy
vm2015_merge_filename <- paste(amphibdir_data_out,"vm2015_merge.csv", sep="")
write.csv(vm2015_merge, file=vm2015_merge_filename)
```

The data set's dimensions are:

```
## [1] 196  12
```

Van Meter et al. 2016

From the initial 11 columns, 4 columns were dropped and consolidated into 1, and 4 columns were added.

```
# add sample_id
vm2016$sample_id <- paste(vm2016$Day, vm2016$Row, vm2016$Column, sep="_")
vm2016_subset2 <- subset(vm2016, select=c(-Day,-Row, -Column, -Total))
# add additional columns
species <- c(rep("American toad",nrow(vm2016_subset2)))
application <- c(rep("Indirect",nrow(vm2016_subset2)))
exp_duration<- (rep(8,nrow(vm2016_subset2)))
source <- c(rep("rvm2016",nrow(vm2016_subset2)))
vm2016_subset3 <- cbind(vm2016_subset2, species, application, exp_duration, source)
```

Application rates for several pesticides were inserted. There were 108 observations with decay products that were not sprayed; these observations were dropped so as to only include the parents in the cleaned data set. There were 60 observations with atrazine, fipronil, or triadimefon that were dropped because they do not include metabolites in total.

```
# assign values to application rate
#unique(vm2016_subset3$Pesticide)
vm2016_subset3$app_rate_g_cm2[vm2016_subset3$Pesticide=="ATZTOT"] <- 22.9e-6
vm2016_subset3$app_rate_g_cm2[vm2016_subset3$Pesticide=="Imid"] <- 5.7e-6
vm2016_subset3$app_rate_g_cm2[vm2016_subset3$Pesticide=="FipTOT"] <- 1.1e-6
vm2016_subset3$app_rate_g_cm2[vm2016_subset3$Pesticide=="TNDTOT"] <- 2.7e-6
vm2016_subset3$app_rate_g_cm2[vm2016_subset3$Pesticide=="Pendi"] <- 69.8e-6
# drop decay products that were not sprayed, keeping only parents
rows_to_drop <- which(vm2016_subset3$Parent == 0)
vm2016_subset4 <- vm2016_subset3[-rows_to_drop,]
# drop ATZ, Fip, TDN since do not include metabolites in total
chems_to_drop <- c("ATZ","Fip","TDN")
vm2016_subset5 <- vm2016_subset4[!(vm2016_subset4$Pesticide %in% chems_to_drop),]
# drop parent field
drop_cols <- c("Parent")
vm2016_subset6 <- vm2016_subset5[,!(names(vm2016_subset5) %in% drop_cols)]
```

Several column names were standardized and all columns were ordered for ease of merging with the combined data set.

```
# standardize column names
colnames(vm2016_subset6)
```

```
##  [1] "Pesticide"      "SoilType"       "BodyBurden"     "Soil"
##  [5] "Weight"         "Formulation"    "sample_id"      "species"
##  [9] "application"    "exp_duration"   "source"         "app_rate_g_cm2"
```

```
colnames(vm2016_subset6)[which(colnames(vm2016_subset6)=="Pesticide")]<-"chemical"
colnames(vm2016_subset6)[which(colnames(vm2016_subset6)=="SoilType")]<-"soil_type"
colnames(vm2016_subset6)[which(colnames(vm2016_subset6)=="BodyBurden")]<-"tissue_conc_ugg"
colnames(vm2016_subset6)[which(colnames(vm2016_subset6)=="Soil")]<-"soil_conc_ugg"
colnames(vm2016_subset6)[which(colnames(vm2016_subset6)=="Weight")]<-"body_weight_g"
colnames(vm2016_subset6)[which(colnames(vm2016_subset6)=="Formulation")]<-"formulation"
```

```
# alter chemical name
vm2016_subset6$chemical <- as.character(vm2016_subset6$chemical)
vm2016_subset6$chemical[vm2016_subset6$chemical=="Imid"] <- "imidacloprid"

# reorder columns alphabetically to help with merge
colnames(vm2016_subset6)
```

```
##  [1] "chemical"        "soil_type"       "tissue_conc_ugg" "soil_conc_ugg"
##  [5] "body_weight_g"   "formulation"     "sample_id"       "species"
##  [9] "application"     "exp_duration"    "source"          "app_rate_g_cm2"
```

```
vm2016_merge <- vm2016_subset6[,order(names(vm2016_subset6))]
colnames(vm2016_merge)
```

```
##  [1] "app_rate_g_cm2"  "application"     "body_weight_g"   "chemical"
##  [5] "exp_duration"    "formulation"     "sample_id"       "soil_conc_ugg"
##  [9] "soil_type"       "source"          "species"         "tissue_conc_ugg"
```

```
# write a local copy
vm2016_merge_filename <- paste(amphibdir_data_out,"vm2016_merge.csv", sep="")
write.csv(vm2016_merge, file=vm2016_merge_filename)
```

The updated dimensions are:

```
## [1] 96 12
```

The Van Meter et al. 2014/2015 and Van Meter et al. 2016 data sets were combined.

The combined data set's updated dimensions are:

```
## [1] 292  12
```

Glinski et al. 2018a (Dehydration)

The metabolite products were dropped from the data set; 600 rows from the initial 1494 rows were retained.

```
# drop metabolite products
parent_keepers <- which(as.vector(dag2016_dehy0$parent) == as.vector(dag2016_dehy0$analyte))
dag2016_dehy1 <- dag2016_dehy0[parent_keepers,]
```

Several column names were altered for standarization across the data set, and 7 columns were added for standarization.

```
## time is length of dehydration
#colnames(dag2016_dehy1)[which(colnames(dag2016_dehy1)=="time")]<-"exp_duration"

# standardize column names
colnames(dag2016_dehy1)[which(colnames(dag2016_dehy1)=="analyte")]<-"chemical"
colnames(dag2016_dehy1)[which(colnames(dag2016_dehy1)=="conc")]<-"tissue_conc_ugg"
colnames(dag2016_dehy1)[which(colnames(dag2016_dehy1)=="ID")]<-"sample_id"
```

```r
colnames(dag2016_dehy1)[which(colnames(dag2016_dehy1)=="weight")]<-"body_weight_g"

# add additional columns
exp_duration <- c(rep(8,nrow(dag2016_dehy1)))
soil_type <- c(rep("PLE",nrow(dag2016_dehy1)))
application <- c(rep("Indirect",nrow(dag2016_dehy1)))
formulation <- (rep(0,nrow(dag2016_dehy1)))
app_rate_g_cm2 <- (rep(0,nrow(dag2016_dehy1)))
soil_conc_ugg <- (rep(0,nrow(dag2016_dehy1)))
source <- c(rep("dag_dehydration",nrow(dag2016_dehy1)))
dag2016_dehy2 <- cbind(dag2016_dehy1, formulation, soil_type, application,
                       app_rate_g_cm2, exp_duration, soil_conc_ugg, source)
```

The updated dimensions are:

```
## [1] 600  15
```

Multiple soil concentration observations were given the same ID. Until a many-to-one merge of soil concentrations could be executed, 300 rows were temporarily dropped. There were also 3 columns dropped.

```r
# drop the soil until we can do a many-to-one merge of soil concentrations
# drop decay products that were not sprayed, keeping only parents
rows_to_drop <- which(dag2016_dehy2$matrix == 'soil')
dag2016_dehy3 <- dag2016_dehy2[-rows_to_drop,]
# parent, time and matrix columns delete
drop_cols <- c("parent","time","matrix")
dag2016_dehy4 <- dag2016_dehy3[,!(names(dag2016_dehy3) %in% drop_cols)]
```

The updated dimensions are:

```
## [1] 300  12
```

The application rate values were inserted, the temporarily dropped soil concentrations were updated to the current data set, and the species names were standardized.

```r
# fill in application rates
#unique(dag2016_dehy4$chemical)
update_atrazine <- which(dag2016_dehy4$chemical == 'atrazine')
dag2016_dehy4$app_rate_g_cm2[update_atrazine] <- 0.00002395 # atrazine g/cm2
update_chloro <- which(dag2016_dehy4$chemical == 'chloro+d')
dag2016_dehy4$app_rate_g_cm2[update_chloro] <-  0.0000443 # chloro g/cm2
update_metol <- which(dag2016_dehy4$chemical == 'metol')
dag2016_dehy4$app_rate_g_cm2[update_metol] <-  0.00003101 # metol g/cm2
update_tdn <- which(dag2016_dehy4$chemical == 'tdn')
dag2016_dehy4$app_rate_g_cm2[update_tdn] <- 0.00000291 # tdn g/cm2
update_imid <- which(dag2016_dehy4$chemical == 'imid')
dag2016_dehy4$app_rate_g_cm2[update_imid] <- 0.00000539 # imid g/cm2

# add back in soil concentrations (in already-made soil_conc_ugg column)
dag2016_soil <- dag2016_dehy2[rows_to_drop,]
dag2016_dehy4$soil_conc_ugg <- dag2016_soil$tissue_conc_ugg
```

```
# rename species names, according to standardized names
dag2016_dehy4$species <- as.character(dag2016_dehy4$species)
dag2016_dehy4$species[dag2016_dehy4$species == "LF"] <- "Leopard frog"
dag2016_dehy4$species[dag2016_dehy4$species == "BA"] <- "Fowlers toad"
dag2016_dehy4$species <- as.factor(dag2016_dehy4$species)
```

The dimensions are:

```
## [1] 300  12
```

The Glinkski et al. 2018a (Dehydration) was combined with the previously merged data sets.

The combined data set's updated dimensions are:

```
## [1] 592  12
```

Henson-Ramsey 2008

The Henson-Ramsey 2008 data set did not require any additional data cleaning. It was combined with the previously merged data sets.

The combined data set's updated dimensions are:

```
## [1] 601  12
```

Glinski et al. 2018b (Metabolites)

Apart from standardizing the species name, the Glinski et al. 2018b (Metabolites) data set did not require any additional data cleaning. It was combined with the previously merged data sets.

```
# rename species names, according to standardized names
dag2016_metabolite_merge$species <- as.character(dag2016_metabolite_merge$species)
dag2016_metabolite_merge$species[dag2016_metabolite_merge$species == "Anaxyrus_fowleri"] <- "Fowlers toa
dag2016_metabolite_merge$species <- as.factor(dag2016_metabolite_merge$species)
```

The combined data set's updated dimensions are:

```
## [1] 661  12
```

Glinski et al. 2019 (Biomarkers)

Five columns were dropped from the original biomarkers data set and the names of two columns were standardized.

```
# drop columns
drop_cols <- c("met", "tdt", "bif", "rate", "group")
dag_biomarker_subset <- dag_biomarker[, !(names(dag_biomarker) %in% drop_cols)]

# standardize column names
colnames(dag_biomarker_subset)[which(colnames(dag_biomarker_subset)=="conc")]<-"tissue_conc_ugg"
colnames(dag_biomarker_subset)[which(colnames(dag_biomarker_subset)=="frog.weight")]<-"body_weight_g"
```

The updated column names and dimensions are:

```
## [1] "body_weight_g"   "sample_id"       "pesticide"       "tissue_conc_ugg"
```

```
## [1] 192   4
```

The application rates and soil concentrations were not included in the original biomarkers data set. Both are included in the following data set:

Data Set Dimensions, Column Names, and Summary:

```
## [1] 136  15
```

```
##  [1] "frog.weight"  "SAMPLE"       "Met"          "TDN"          "TDL"
##  [6] "BIF"          "soil.weight"  "Met.soil"     "TDN.soil"     "TDL.soil"
## [11] "BIF.soil"     "Rate"         "app.rate.met" "app.rate.tdn" "app.rate.bif"
```

```
##   frog.weight       SAMPLE              Met               TDN
##  Min.   :1.012   Length:136         Min.   : 0.0000   Min.   :0.00000
##  1st Qu.:2.749   Class :character   1st Qu.: 0.0000   1st Qu.:0.00000
##  Median :3.164   Mode  :character   Median : 0.0000   Median :0.00000
##  Mean   :3.302                      Mean   : 0.9123   Mean   :0.06927
##  3rd Qu.:3.762                      3rd Qu.: 0.4298   3rd Qu.:0.07447
##  Max.   :6.784                      Max.   :19.8798   Max.   :0.55921
##       TDL                BIF          soil.weight        Met.soil
##  Min.   :0.00000   Min.   :0.0000   Min.   : 4.476   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.: 6.731   1st Qu.:0.000
##  Median :0.00000   Median :0.0000   Median : 7.772   Median :0.000
##  Mean   :0.02259   Mean   :0.1276   Mean   : 8.043   Mean   :1.605
##  3rd Qu.:0.01770   3rd Qu.:0.1299   3rd Qu.: 9.050   3rd Qu.:2.265
##  Max.   :0.30815   Max.   :1.0271   Max.   :13.571   Max.   :6.758
##     TDN.soil          TDL.soil           BIF.soil          Rate
##  Min.   :0.0000   Min.   :0.000000   Min.   :0.0000   Length:136
##  1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.:0.0000   Class :character
##  Median :0.0000   Median :0.000000   Median :0.0000   Mode  :character
##  Mean   :0.7168   Mean   :0.010160   Mean   :0.7417
##  3rd Qu.:0.5312   3rd Qu.:0.007463   3rd Qu.:1.1472
##  Max.   :3.6300   Max.   :0.061563   Max.   :5.2658
##   app.rate.met     app.rate.tdn     app.rate.bif
##  Min.   : 0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median : 0.000   Median :0.0000   Median :0.0000
##  Mean   :14.263   Mean   :1.3389   Mean   :1.6070
##  3rd Qu.: 5.511   3rd Qu.:0.5173   3rd Qu.:0.6209
##  Max.   :55.106   Max.   :5.1730   Max.   :6.2090
```

The application rates were converted from mg to g/cm^2.

```
dag_biomarker2_update <- replace.value(dag_biomarker2, "app.rate.met", from=55.106, to=3.062e-5, verbos
dag_biomarker2_update <- replace.value(dag_biomarker2_update, "app.rate.met", from=5.5106, to=3.062e-6,
dag_biomarker2_update <- replace.value(dag_biomarker2_update, "app.rate.tdn", from=5.173, to=2.87e-6, v
dag_biomarker2_update <- replace.value(dag_biomarker2_update, "app.rate.tdn", from=.5173, to=2.87e-7, v
dag_biomarker2_update <- replace.value(dag_biomarker2_update, "app.rate.bif", from=6.209, to=3.45e-6, v
dag_biomarker2_update <- replace.value(dag_biomarker2_update, "app.rate.bif", from=.6209, to=3.45e-7, v
```

A one-to-one merge was conducted based on the unique sample id for each measured pesticide (either bifenthrin, metolachlor, or triadimefon) to join the original biomarkers data set and the data set containing the application rates and soil concentrations. Vectors containing the application rates and soil concentrations were joined to the original data set.

```
# bif extraction
dag_biomarker_subset_bif <- dag_biomarker_subset[dag_biomarker_subset$pesticide == "bif", ]
dag_biomarker2_subset_bif <- dag_biomarker2_update[dag_biomarker2_update$BIF != 0, ]

dag_biomarker_bif_merge <- merge(x = dag_biomarker_subset_bif, y = dag_biomarker2_subset_bif,
                                 by.x = "sample_id", by.y = "SAMPLE", all.x = TRUE)

# met extraction
dag_biomarker_subset_met <- dag_biomarker_subset[dag_biomarker_subset$pesticide == "met", ]
dag_biomarker2_subset_met <- dag_biomarker2_update[dag_biomarker2_update$Met != 0, ]

dag_biomarker_met_merge <- merge(x = dag_biomarker_subset_met, y = dag_biomarker2_subset_met,
                                 by.x = "sample_id", by.y = "SAMPLE", all.x = TRUE)

# tdt extraction
dag_biomarker_subset_tdt <- dag_biomarker_subset[dag_biomarker_subset$pesticide == "tdt", ]
dag_biomarker2_subset_tdt <- dag_biomarker2_update[dag_biomarker2_update$TDN != 0, ]

dag_biomarker_tdt_merge <- merge(x = dag_biomarker_subset_tdt, y = dag_biomarker2_subset_tdt,
                                 by.x = "sample_id", by.y = "SAMPLE", all.x = TRUE)


# combine bif, met, and tdt
app_bind_bmt <- c(dag_biomarker_bif_merge[,"app.rate.bif"],
                  dag_biomarker_met_merge[,"app.rate.met"], dag_biomarker_tdt_merge[,"app.rate.tdn"])

soil_bind_bmt <- c(dag_biomarker_bif_merge[,"BIF.soil"],
                   dag_biomarker_met_merge[,"Met.soil"], dag_biomarker_tdt_merge[,"TDN.soil"])


# join app and soil vectors to data set
dag_biomarker_subset2 <- dag_biomarker_subset[order(dag_biomarker_subset[, 3]),]
rownames(dag_biomarker_subset2) <- seq(length=nrow(dag_biomarker_subset2))

dag_biomarker_subset3 <- cbind(dag_biomarker_subset2, app_bind_bmt, soil_bind_bmt)

# standardize column names
colnames(dag_biomarker_subset3)[which(colnames(dag_biomarker_subset3)=="app_bind_bmt")]<-"app_rate_g_cm
colnames(dag_biomarker_subset3)[which(colnames(dag_biomarker_subset3)=="soil_bind_bmt")]<-"soil_conc_ug
```

The updated column names and dimensions are:

```
## [1] "body_weight_g"   "sample_id"       "pesticide"       "tissue_conc_ugg"
## [5] "app_rate_g_cm2"  "soil_conc_ugg"

## [1] 192   6
```

New columns were created for standarization, the columns were ordered alphabetically, and a local copy was stored.

```
# create new columns
application <- c(rep("soil", nrow(dag_biomarker_subset3)))
exp_duration <- c(rep(8, nrow(dag_biomarker_subset3)))
formulation <- c(rep(0, nrow(dag_biomarker_subset3)))
soil_type <- c(rep(NA, nrow(dag_biomarker_subset3)))
source <- c(rep("dag_biomarker", nrow(dag_biomarker_subset3)))
species <- c(rep("Leopard frog", nrow(dag_biomarker_subset3)))

# combine columns
dag_biomarker_subset4 <- cbind(dag_biomarker_subset3, application, exp_duration,
                               formulation, soil_type, source, species)

# standardize pesticide column
dag_biomarker_subset4$pesticide <- as.character(dag_biomarker_subset4$pesticide)
dag_biomarker_subset4$pesticide[dag_biomarker_subset4$pesticide == "bif"] <- "Bifenthrin"
dag_biomarker_subset4$pesticide[dag_biomarker_subset4$pesticide == "met"] <- "Metolachlor"
dag_biomarker_subset4$pesticide[dag_biomarker_subset4$pesticide == "tdt"] <- "Triadimefon"

colnames(dag_biomarker_subset4)[which(colnames(dag_biomarker_subset4)=="pesticide")]<-"chemical"

# unite function for sample id and chemical
dag_biomarker_subset5 <- unite(data = dag_biomarker_subset4, col = "sample_id", "sample_id", "chemical"

# order columns in abc for merge
dag_biomarker_merge <- dag_biomarker_subset5[ ,order(names(dag_biomarker_subset5))]
```

The updated column names and dimensions are:

```
##  [1] "app_rate_g_cm2"  "application"     "body_weight_g"   "chemical"
##  [5] "exp_duration"    "formulation"     "sample_id"       "soil_conc_ugg"
##  [9] "soil_type"       "source"          "species"         "tissue_conc_ugg"
```

```
## [1] 192  12
```

The Glinski et al. 2019 (Biomarkers) was combined with the previously merged data sets.

The combined data set's updated dimensions are:

```
## [1] 853  12
```

Van Meter et al. 2018 (Multiple Pesticides Study)

The Van Meter et al. 2018 (Multiple Pesticides Study) data set did not require any additional data cleaning. It was combined with the previously merged data sets.

The combined data set's updated dimensions are:

```
## [1] 990  12
```

Glinski et al. 2020 (Dermal Routes)

The dermal routes data set did not include the body weights for the measured amphibians. These weights were included in a separate data set:

Data Set Dimensions, Column Names, and Summary:

```
## [1] 48  2
```

```
## [1] "Weight_g" "Sample"
```

```
##     Weight_g         Sample
##  Min.   :0.9555   Length:48
##  1st Qu.:1.4204   Class :character
##  Median :1.7817   Mode  :character
##  Mean   :1.7784
##  3rd Qu.:2.1319
##  Max.   :2.8197
```

A one-to-many merge was employed to merge the dermal routes data set and the weights data set based on the Sample ID. Only rows where the Matrix is "Amphibian" have a body weight; all other rows are NA.

```
# merge (one-to-many) dermal routes data with weights data, based on Sample ID
dermal_routes_subset2 <- dermal_routes[order(dermal_routes$Sample.ID), ]
weights_2 <- weights[order(weights$Sample),]

dermal_routes_subset3 <- merge(dermal_routes_subset2, weights_2,
                               by.x = "Sample.ID", by.y = "Sample", all.x = TRUE, all.y = TRUE)
```

The updated dimensions are:

```
## [1] 192   6
```

The soil concentrations, where the Media and Matrix are both "Soil," was subset from the data set to be used later in the data cleaning process. These soil concentrations (currently listed in the "Concentration" column) will be used for the soil_conc_ugg column in the cleaned data set.

```
# subset soil to be used later for soil concentration column (will use "Concentration" column)
soil_subset <- dermal_routes_subset2[dermal_routes_subset2$Media == "Soil", ]
soil_subset2 <- soil_subset[soil_subset$Matrix == "Soil",]
```

The dimensions of this soil subset are:

```
## [1] 48  5
```

Referring back to the main dermal routes data set: we are only interested in the pesticide exposures on amphibians while in soil. These rows were subset.

```
# want Media == soil because interested in dermal exposure in soil
dermal_routes_subset4 <- dermal_routes_subset3[dermal_routes_subset3$Media == "Soil",]
#sum(dermal_routes_subset3$Media == "Soil") # == 96
#dim(dermal_routes_subset4) # == 96 x 6

# want Matrix == Amphibian because interested in amphib exposure
dermal_routes_subset5 <- dermal_routes_subset4[dermal_routes_subset4$Matrix == "Amphibian", ]
#sum(dermal_routes_subset4$Matrix == "Amphibian") # == 48
#dim(dermal_routes_subset5) # == 48 x 6
```

The updated dimensions are:

```
## [1] 48  6
```

The soil concentrations were appended to the main dermal routes data set.

```
# add in soil concentration column, previously subset
# order by Sample.ID, then by Analyte name to match up rows for the two data sets
dermal_routes_subset6 <- dermal_routes_subset5[order(dermal_routes_subset5[,1],
                                                     dermal_routes_subset5[,2]),]
soil_subset3 <- soil_subset2[order(soil_subset2[,1], soil_subset2[,2]),]

#dim(dermal_routes_subset6) # == 48 x 6
#dim(soil_subset3) # == 48 x 5

dermal_routes_subset7 <- cbind(dermal_routes_subset6, soil_subset3$Concentration)
```

The updated dimensions are:

```
## [1] 48  7
```

The metabolites were dropped from the data set. Additionally, several new columns were created for standarization, existing columns were standardized according to the naming conventions of the collated data set, and unneeded columns were dropped. Columns were ordered alphabetically for ease of merging.

```
# drop metabolites
rows_to_drop <- c("4-OH", "CPO", "TFSa")
dermal_routes_subset8 <- dermal_routes_subset7[!(dermal_routes_subset7$Analyte %in% rows_to_drop),]

# create new columns
app_rate_g_cm2 <- c(rep(NA, nrow(dermal_routes_subset8)))
application <- c(rep("soil", nrow(dermal_routes_subset8)))
exp_duration <- c(rep(8, nrow(dermal_routes_subset8)))
formulation <- c(rep(0, nrow(dermal_routes_subset8)))
soil_type <- c(rep("OLS", nrow(dermal_routes_subset8)))
source <- c(rep("dag_dermal_routes", nrow(dermal_routes_subset8)))
species <- c(rep("Leopard frog", nrow(dermal_routes_subset8)))

# alter existing column names
colnames(dermal_routes_subset8)
```

```
## [1] "Sample.ID"                  "Analyte"
## [3] "Media"                      "Matrix"
## [5] "Concentration"              "Weight_g"
## [7] "soil_subset3$Concentration"
```

```
colnames(dermal_routes_subset8)[which(colnames(dermal_routes_subset8)=="Analyte")]<-"chemical"
colnames(dermal_routes_subset8)[which(colnames(dermal_routes_subset8)=="Sample.ID")]<-"sample_id"
colnames(dermal_routes_subset8)[which(colnames(dermal_routes_subset8)=="Concentration")]<-"tissue_conc_
colnames(dermal_routes_subset8)[which(colnames(dermal_routes_subset8)=="soil_subset3$Concentration")]<-
colnames(dermal_routes_subset8)[which(colnames(dermal_routes_subset8)=="Weight_g")]<-"body_weight_g"


# combine columns
dermal_routes_subset9 <- cbind(dermal_routes_subset8, app_rate_g_cm2, application, exp_duration,
                               formulation, soil_type, source, species)
names(dermal_routes_subset9)
```

```
##  [1] "sample_id"       "chemical"        "Media"           "Matrix"
##  [5] "tissue_conc_ugg" "body_weight_g"   "soil_conc_ugg"   "app_rate_g_cm2"
##  [9] "application"     "exp_duration"    "formulation"     "soil_type"
## [13] "source"          "species"
```

```
# drop columns
cols_to_drop <- c("Matrix", "Media")
dermal_routes_subset10 <- dermal_routes_subset9[, !(names(dermal_routes_subset9) %in% cols_to_drop)]

# insert application rates
dermal_routes_subset10$chemical <- as.character(dermal_routes_subset10$chemical)
unique(dermal_routes_subset10$chemical)
```

```
## [1] "BIF" "CPF" "TFS"
```

```
dermal_routes_subset10$app_rate_g_cm2[dermal_routes_subset10$chemical =="BIF"] <- 2.8889e-7 #bifenthrin
dermal_routes_subset10$app_rate_g_cm2[dermal_routes_subset10$chemical =="CPF"] <- 3.1111e-7 #chlorpyrif
dermal_routes_subset10$app_rate_g_cm2[dermal_routes_subset10$chemical =="TFS"] <- 3.0222e-7 #trifloxyst

summary(dermal_routes_subset10$app_rate_g_cm2)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 2.889e-07 2.889e-07 3.022e-07 3.007e-07 3.111e-07 3.111e-07
```

```
# order columns in abc for merge
dermal_routes_merge <- dermal_routes_subset10[ ,order(names(dermal_routes_subset10))]
```

The updated column names and dimensions are:

```
## [1] 24 12
```

```
##  [1] "app_rate_g_cm2"  "application"     "body_weight_g"   "chemical"
##  [5] "exp_duration"    "formulation"     "sample_id"       "soil_conc_ugg"
##  [9] "soil_type"       "source"          "species"         "tissue_conc_ugg"
```

A local copy was saved, and the data set was combined with the collated data set.

The combined data set's updated dimensions are:

```
## [1] 1014   12
```

Final Product

Minor alterations were made to the final collated data set to standardize names of the application types and chemicals.

```
amphib_dermal_collated <- combined_data6

colnames(amphib_dermal_collated)
```

```
##  [1] "app_rate_g_cm2"  "application"     "body_weight_g"   "chemical"
##  [5] "exp_duration"    "formulation"     "sample_id"       "soil_conc_ugg"
##  [9] "soil_type"       "source"          "species"         "tissue_conc_ugg"
```

```
# check to see if everything ok
summary(amphib_dermal_collated$app_rate_g_cm2)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 2.870e-07 2.790e-06 5.700e-06 1.616e-05 2.395e-05 6.980e-05
```

```
summary(amphib_dermal_collated$body_weight_g)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1879  1.3043  2.1247  3.3800  3.0412 50.9200
```

```
summary(amphib_dermal_collated$exp_duration)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   8.000   8.000   8.876   8.000  48.000
```

```
summary(amphib_dermal_collated$soil_conc_ugg) # 206 NAs
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##   0.1125   2.0709   5.2459  14.3042  15.3781 238.1502      206
```

```
summary(amphib_dermal_collated$tissue_conc_ugg)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  0.00054  0.16908  0.52573  2.51415  2.06812 72.62672
```

```r
# standardize application levels
amphib_dermal_collated$application <- tolower(amphib_dermal_collated$application)
amphib_dermal_collated$application <- as.factor(amphib_dermal_collated$application)

# standardize chemical levels
amphib_dermal_collated$chemical <- as.character(amphib_dermal_collated$chemical)

amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "fip"] <- "fipronil"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "BIF"] <- "bifenthrin"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "MET"] <- "metolachlor"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "MAT"] <- "malathion"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "ATZT"] <- "atrazine"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "PROPT"] <- "propiconazole"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "metol"] <- "metolachlor"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "tdn"] <- "triadimefon"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "imid"] <- "imidacloprid"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "chloro+d"] <- "chlorothalonil"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "CPF"] <- "chlorpyrifos"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "TFS"] <- "trifloxystrobin"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "FipTOT"] <- "fipronil"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "ATZTOT"] <- "atrazine"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "TNDTOT"] <- "triadimefon"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "Pendi"] <- "pendimethalin"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "Total Atrazine"] <- "atrazine"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "Total Fipronil"] <- "fipronil"
amphib_dermal_collated$chemical[amphib_dermal_collated$chemical == "Total Triadimefon"] <- "triadimefon


amphib_dermal_collated$chemical <- tolower(amphib_dermal_collated$chemical)
amphib_dermal_collated$chemical <- as.factor(amphib_dermal_collated$chemical)

# write out file
amphib_dermal_collated_filename <- paste(amphibdir_data_out,"amphib_dermal_collated.csv", sep="")
write.csv(amphib_dermal_collated, file=amphib_dermal_collated_filename)
```

Column Names

```
##  [1] "app_rate_g_cm2"  "application"     "body_weight_g"   "chemical"
##  [5] "exp_duration"    "formulation"     "sample_id"       "soil_conc_ugg"
##  [9] "soil_type"       "source"          "species"         "tissue_conc_ugg"
```

Dimensions

```
## [1] 1014   12
```

Variable Summaries

```
##  app_rate_g_cm2         application  body_weight_g             chemical
##  Min.   :2.870e-07   indirect :396   Min.   : 0.1879   triadimefon :223
##  1st Qu.:2.790e-06   overspray: 45   1st Qu.: 1.3043   atrazine    :191
##  Median :5.700e-06   soil     :573   Median : 2.1247   metolachlor :154
```

```
##  Mean   :1.616e-05                    Mean   : 3.3800   imidacloprid: 78
##  3rd Qu.:2.395e-05                    3rd Qu.: 3.0412   bifenthrin  : 72
##  Max.   :6.980e-05                    Max.   :50.9200   fipronil    : 71
##                                                         (Other)     :225
##   exp_duration     formulation       sample_id         soil_conc_ugg
##  Min.   : 2.000   Min.   :0.00000   Length:1014        Min.   :  0.1125
##  1st Qu.: 8.000   1st Qu.:0.00000   Class :character   1st Qu.:  2.0709
##  Median : 8.000   Median :0.00000   Mode  :character   Median :  5.2459
##  Mean   : 8.876   Mean   :0.03582                      Mean   : 14.3042
##  3rd Qu.: 8.000   3rd Qu.:0.00000                      3rd Qu.: 15.3781
##  Max.   :48.000   Max.   :1.00000                      Max.   :238.1502
##                   NA's   :9                            NA's   :206
##   soil_type            source            species          tissue_conc_ugg
##  Length:1014        Length:1014        Length:1014        Min.   : 0.00054
##  Class :character   Class :character   Class :character   1st Qu.: 0.16908
##  Mode  :character   Mode  :character   Mode  :character   Median : 0.52573
##                                                           Mean   : 2.51415
##                                                           3rd Qu.: 2.06812
##                                                           Max.   :72.62672
##
```

Session Information

```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 18363)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] tinytex_0.32     anchors_3.0-8    MASS_7.3-51.6    rgenoud_5.8-3.0
##  [5] stringr_1.4.0    tidyr_1.1.2      dplyr_1.0.5      knitr_1.31
##  [9] kableExtra_1.3.4 reshape2_1.4.4   gridExtra_2.3    ggplot2_3.3.3
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.6        pillar_1.6.0      compiler_4.0.2    plyr_1.8.6
##  [5] tools_4.0.2       digest_0.6.27     viridisLite_0.3.0 evaluate_0.14
##  [9] lifecycle_1.0.0   tibble_3.1.1      gtable_0.3.0      pkgconfig_2.0.3
## [13] rlang_0.4.10      rstudioapi_0.13   DBI_1.1.1         yaml_2.2.1
## [17] xfun_0.24         xml2_1.3.2        httr_1.4.2        withr_2.4.1
## [21] systemfonts_1.0.2 generics_0.1.0    vctrs_0.3.7       webshot_0.5.2
```

```
## [25] grid_4.0.2        tidyselect_1.1.0  svglite_2.0.0     glue_1.4.1
## [29] R6_2.5.0          fansi_0.4.2       rmarkdown_2.6     purrr_0.3.4
## [33] magrittr_2.0.1    scales_1.1.1      ellipsis_0.3.1    htmltools_0.5.1.1
## [37] rvest_0.3.6       assertthat_0.2.1  colorspace_1.4-1  utf8_1.2.1
## [41] stringi_1.5.3     munsell_0.5.0     crayon_1.4.1
```