Application feature improvements in support of human health assessments: optimisations for epidemiology data extraction

Sean Watford, Krista Christensen*, Elizabeth Radke-Farabaugh, Andy Shapiro

Center for Public Health and Environmental Assessment

Office of Research and Development

U.S. Environmental Protection Agency

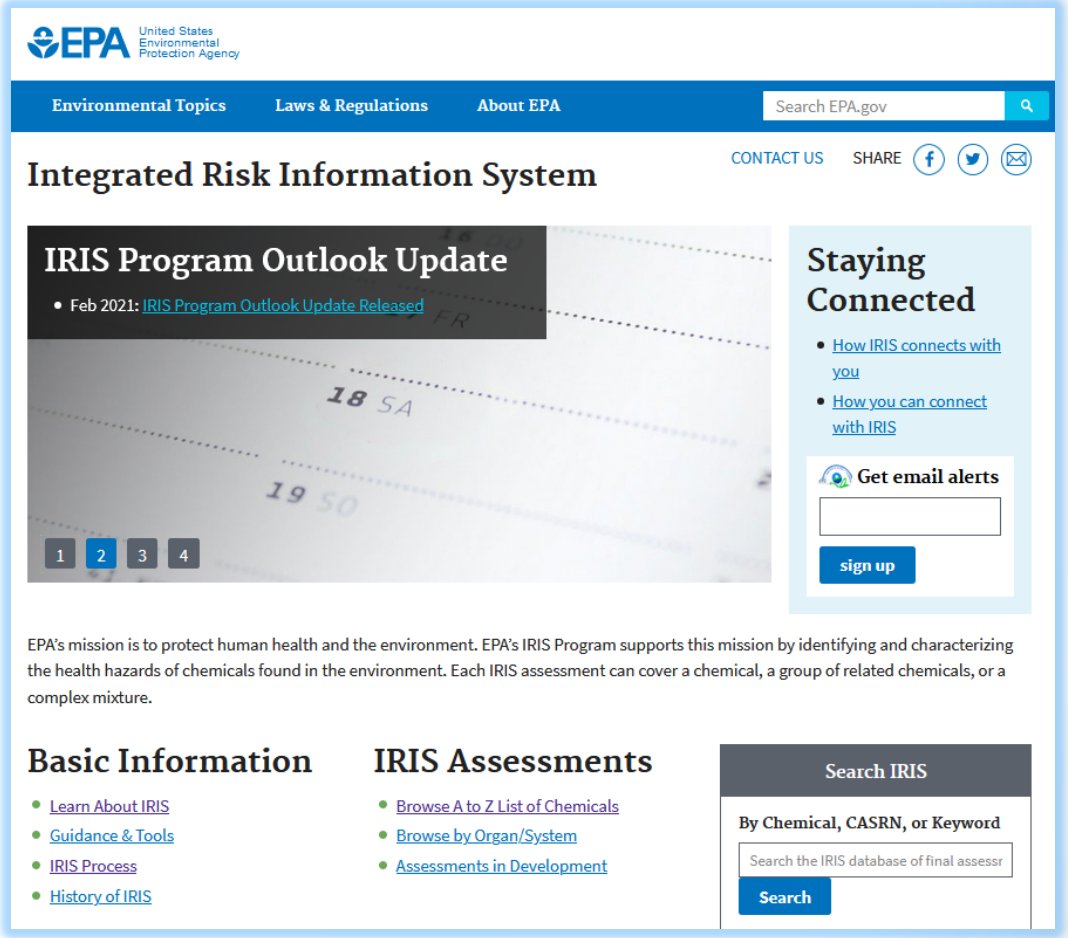
10 November 2023

Disclaimer

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

EPA's Integrated Risk Information System (IRIS)

- Database of health effects information on hundreds of environmental pollutants
- IRIS assessments contribute to decisions across EPA and other health agencies
- Focus is on toxicity due to lifetime exposure
- Provides toxicity values for cancer and noncancer effects
- Have no direct regulatory impact until combined with extent of exposure, cost of cleanup, available technology, and other regulatory options that are the purview of other EPA programs



EPA's Integrated Risk Information System (IRIS)

IRIS assessments contribute to EPA decisions such as:

- Health-based national standards
- Health-based clean-up levels at local sites
- Health-based advisory levels
- Ranking across chemicals
- Information for the general public
- Cost-benefit analyses

IRIS Supports

- **Clean Air Act (CAA)**

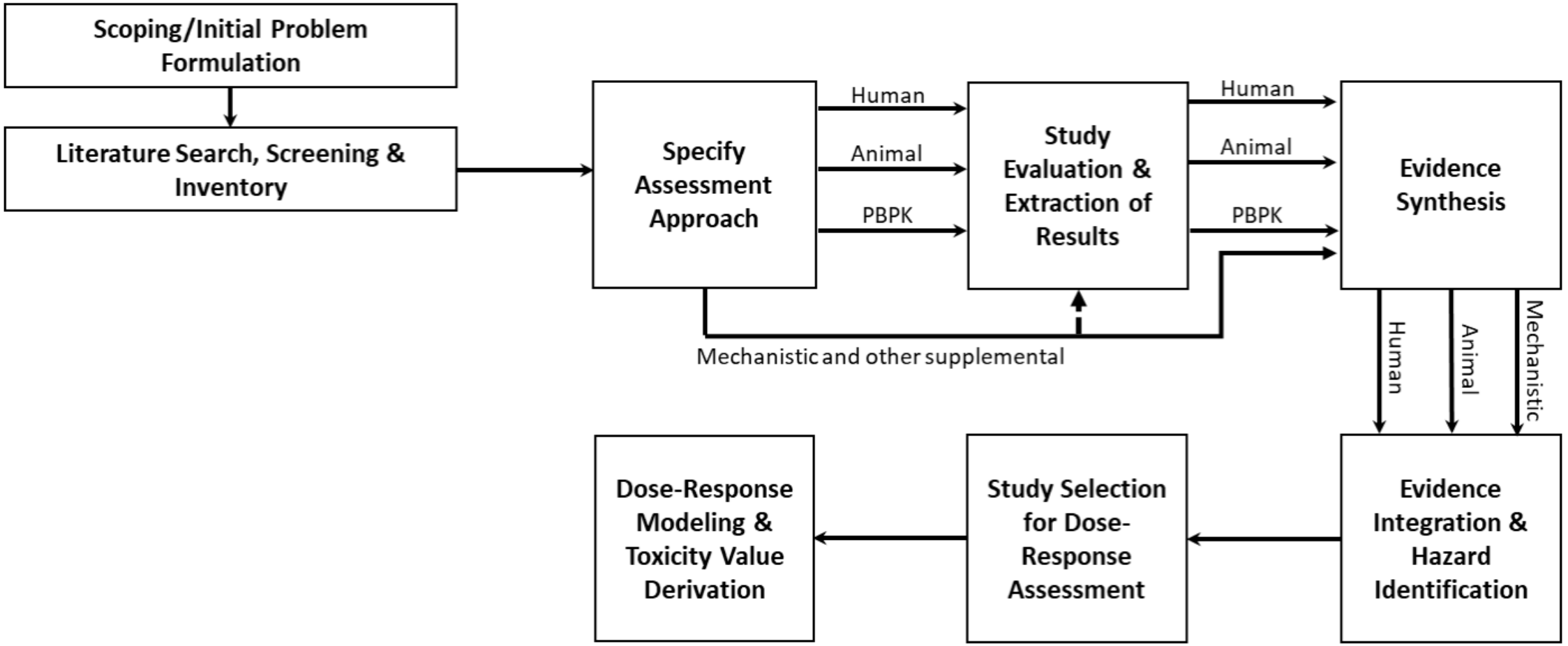
- **Safe Drinking Water Act (SDWA)**
- **Food Quality Protection Act (FQPA)**

- **Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)**
- **Toxic Substances Control Act (TSCA)**
- **Resource Conservation and Recovery Act (RCRA)**

Broad Input to

- **Agency Strategic Goals**
- **Regions and States**
- **Children's Health**
- **Environmental Justice**

IRIS Assessment Development Process



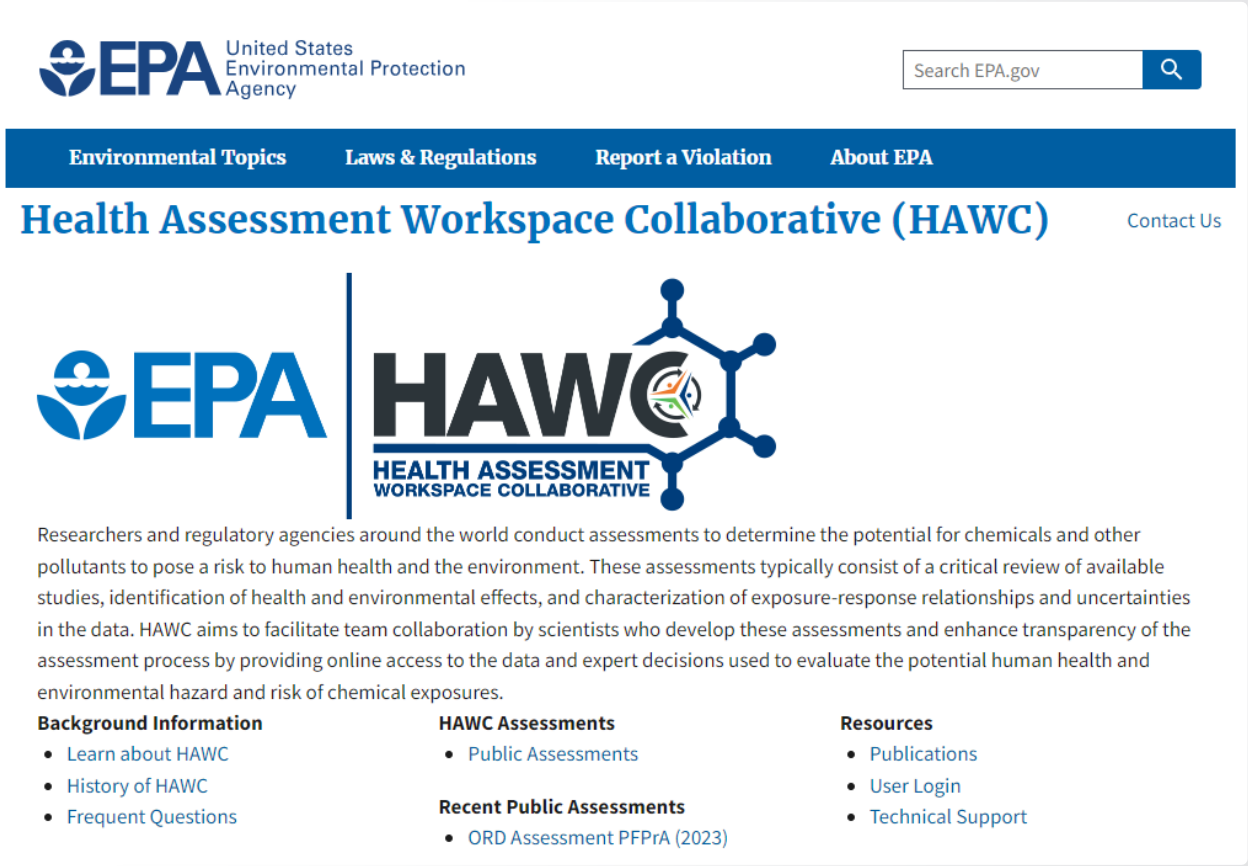
IRIS Assessment Development Process

- These systematic review (SR) methods are resource- and time-intensive, yet increasingly a foundational part of the chemical assessment process
- Software applications and tools can help to:
  - Improve user interfaces and interactions (UI/UX)
  - Standardize data exchange formats
  - Utilize artificial intelligence for (semi-)automation.
- Focus for this presentation is the IRIS program's use of the Health Assessment Workspace Collaborative (HAWC)
  - HAWC has data extraction features for both animal toxicology and epidemiology studies.
  - We will describe recent updates made by the HAWC team in coordination with EPA epidemiologists to update data extraction features including:
    - Updates to UI/UX
    - Increased flexibility to accommodate partial extractions

The Health Assessment Workspace Collaborative (HAWC)

What is HAWC?

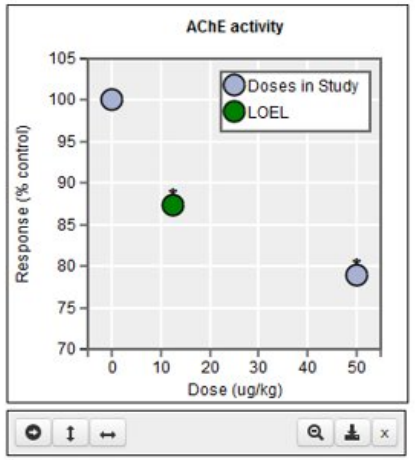
- Web-based content management system for human health data
  - Used by EPA, NTP, WHO/IARC, CalEPA, and others
- Open-source Python application
  - Source code is available at https://github.com/usepa/hawc
  - EPA deployment available for EPA projects and collaborators: hawc.epa.gov
  - Public deployment is available at https://hawcproject.org*
  - Custom, private deployments are possible (MIT licensed)
- Assessments with tiered access
  - Managed read/write access
  - Assessments can be public

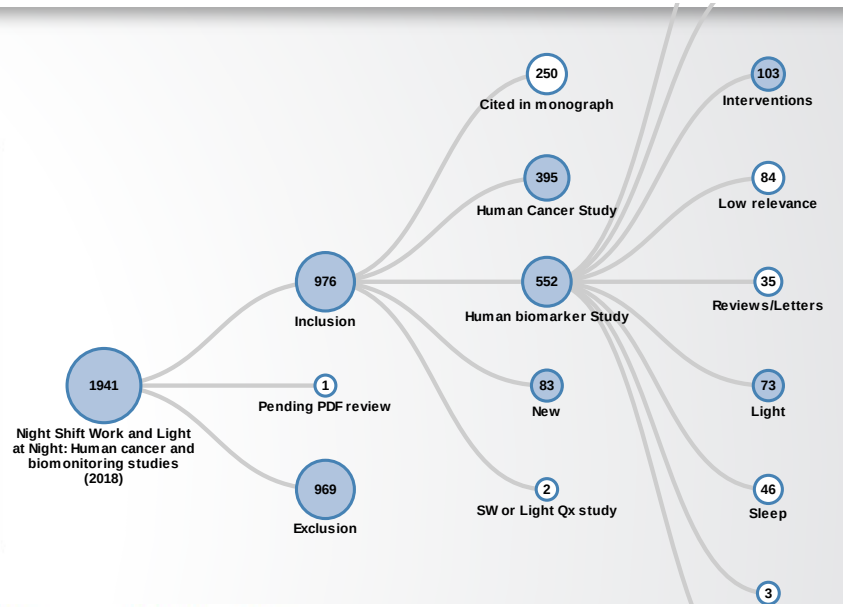


** Mention of or referral links to non-EPA sites does not imply official EPA endorsement of or responsibility for the opinions, ideas, data, or products presented at those locations, or guarantee the validity of the information provided.*

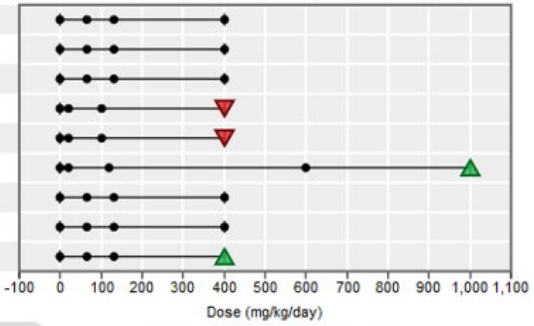
What is HAWC?

- Key modules:
  - Literature screening
  - Risk of bias/study evaluation
  - Animal bioassay data extraction
  - Epidemiology data extraction
  - Interactive summary tables and visualizations
- Interactive "supplemental materials" for reports and data re-use





Endpoint	Reference	Animal Description	Exposure Duration	Route
Clinical Observations	Covance, 2001 (ECHA Summary)	Rat, Sprague-Dawley Crl:Cd Br (♂)	90d	oral gavage
Feed Consumption	Covance, 2001 (ECHA Summary)	Rat, Sprague-Dawley Crl:Cd Br (♂)	90d	oral gavage
		Rat, Sprague-Dawley Crl:Cd Br (♀)	90d	oral gavage
Feed Consumption, Absolute	Argus Research Laboratories Inc, 1998	P0 Rat, Crl:Cd@(Sd)Igs Br Vaf/Plustm (♀, N=23-25)	GD7-17	oral gavage
Feed Consumption, Relative	Argus Research Laboratories Inc, 1998	P0 Rat, Crl:Cd@(Sd)Igs Br Vaf/Plustm (♀, N=23-25)	GD7-17	oral gavage
Mortality	Unnamed Report, 1998 (ECHA Summary)	P0 Rat, Crl:Cd(Sd) Br Vaf/Plus (♀, N=8)	GD7-17	oral gavage
	Covance, 2001 (ECHA Summary)	Rat, Sprague-Dawley Crl:Cd Br (♂)	90d	oral gavage
		Rat, Sprague-Dawley Crl:Cd Br (♀)	90d	oral gavage
Urine-Stained Anogenital Fur	Covance, 2001 (ECHA Summary)	Rat, Sprague-Dawley Crl:Cd Br (♀, N=15)	90d	oral gavage



Participant Selection	Exposure Measurement	Outcome Ascertainment	Confounding	Analysis	Other Sensitivity Concerns	Selective Reporting	Overall study confidence (Epidemiology)
+	–	++	+	+	–	+	–

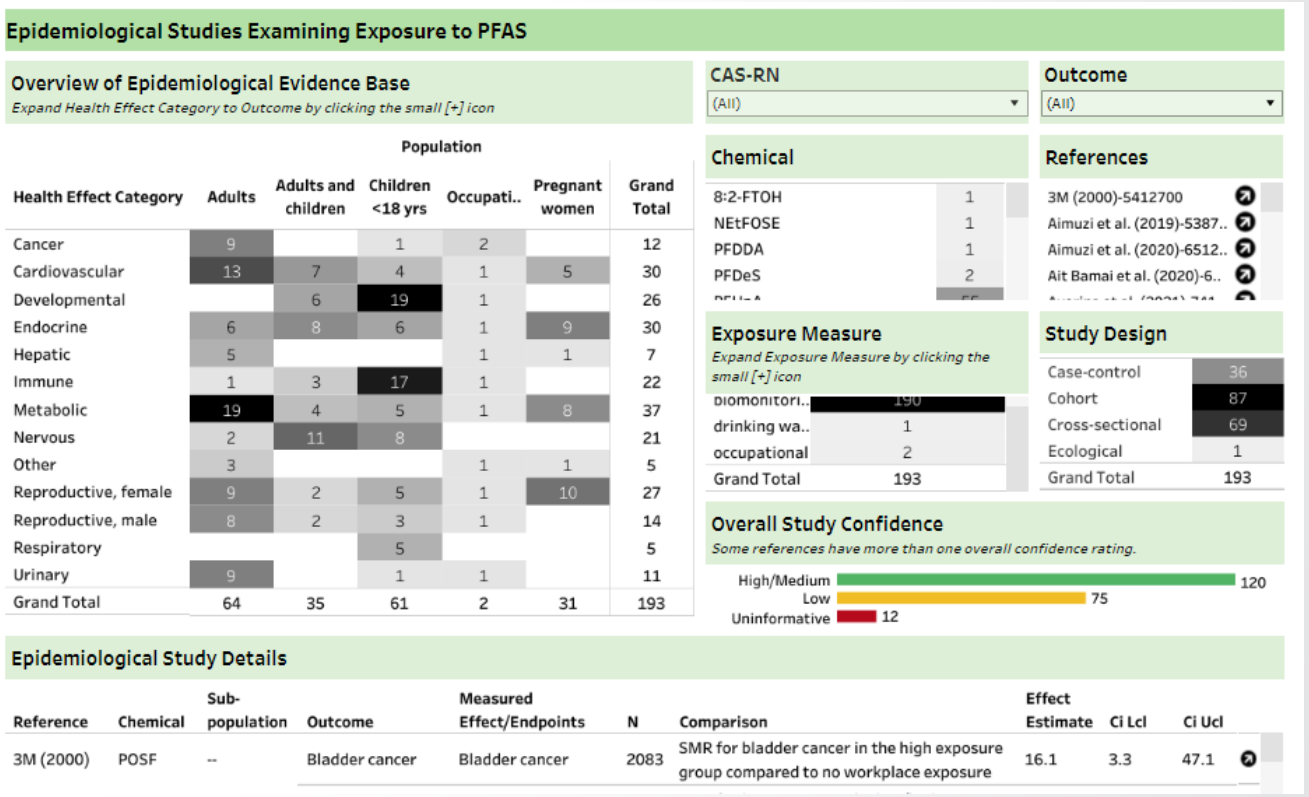
Epidemiology Data Extraction

- Multiple data fields are extracted from epidemiology studies during assessment development
  - Study design features *(e.g., design, population, sample size, time period)*
  - Chemicals *(chemical name and other identifying information like CAS registry number)*
  - Exposures and exposure levels *(e.g., route of exposure, exposure source, distribution of exposure)*
  - Adjustment factors *(e.g., characteristics used for matching or statistical adjustment in multivariate models)*
  - Outcomes *(e.g., assay or instrument used, timing of measurement)*
  - Quantitative results *(e.g., relative risks point and variance estimates)*
- The extraction features are flexible and enable the development and storage of information for multiple purposes: evidence maps, toxicological reviews, visualizations, and interoperability with other tools

Evidence Maps

- High-level summary of the available literature used to inform prioritization, and Scoping and Problem Formulation activities
- Uses tailored criteria to identify and categorize references as potentially relevant to human health risk assessment
- Used to identify key data gap and characterize level of effort and scientific issues to be considered
- Can be used as a tool for identifying needed expertise

[Epidemiology studies examining exposure to PFAS]



Epidemiological Studies Examining Exposure to PFAS

Overview of Epidemiological Evidence Base

Expand Health Effect Category to Outcome by clicking the small [+] icon

Health Effect Category	Population: Adults	Adults and children	Children <18 yrs	Occupati..	Pregnant women	Grand Total
Cancer	9		1	2		12
Cardiovascular	13	7	4	1	5	30
Developmental		6	19	1		26
Endocrine	6	8	6	1	9	30
Hepatic	5			1	1	7
Immune	1	3	17	1		22
Metabolic	19	4	5	1	8	37
Nervous	2	11	8			21
Other	3			1	1	5
Reproductive, female	9	2	5	1	10	27
Reproductive, male	8	2	3	1		14
Respiratory			5			5
Urinary	9		1	1		11
Grand Total	64	35	61	2	31	193

CAS-RN: (All)

Outcome: (All)

Chemical

8:2-FTOH	1
NEtFOSE	1
PFDDA	1
PFDeS	2

References

- 3M (2000)-5412700
- Aimuzi et al. (2019)-5387..
- Aimuzi et al. (2020)-6512..
- Ait Bamai et al. (2020)-6..

Exposure Measure

Expand Exposure Measure by clicking the small [+] icon

biomonitori..	190
drinking wa..	1
occupational	2
Grand Total	193

Study Design

Case-control	36
Cohort	87
Cross-sectional	69
Ecological	1
Grand Total	193

Overall Study Confidence

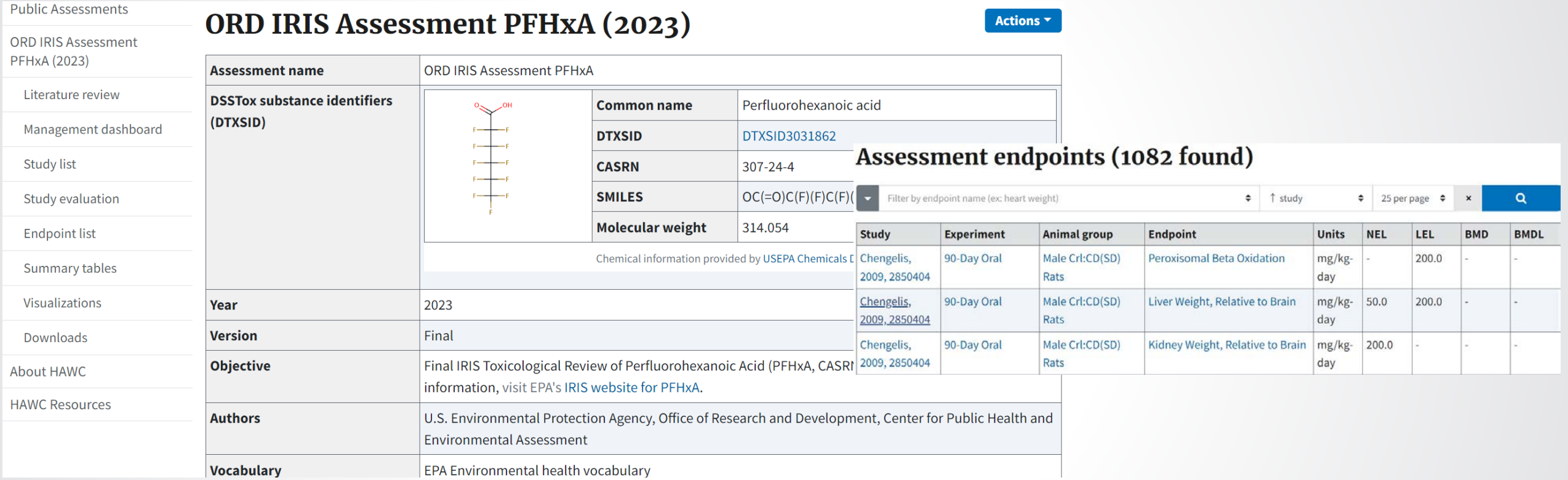
Some references have more than one overall confidence rating.

High/Medium	120
Low	75
Uninformative	12

Epidemiological Study Details

Reference	Chemical	Sub-population	Outcome	Measured Effect/Endpoints	N	Comparison	Effect Estimate	Ci Lcl	Ci Ucl
3M (2000)	POSF	--	Bladder cancer	Bladder cancer	2083	SMR for bladder cancer in the high exposure group compared to no workplace exposure	16.1	3.3	47.1

Toxicological Reviews



Public Assessments

ORD IRIS Assessment PFHxA (2023)

- Literature review
- Management dashboard
- Study list
- Study evaluation
- Endpoint list
- Summary tables
- Visualizations
- Downloads

About HAWC

HAWC Resources

ORD IRIS Assessment PFHxA (2023)

Actions

Assessment name	ORD IRIS Assessment PFHxA
DSSTox substance identifiers (DTXSID)	**Common name**: Perfluorohexanoic acid **DTXSID**: DTXSID3031862 **CASRN**: 307-24-4 **SMILES**: OC(=O)C(F)(F)C(F)(**Molecular weight**: 314.054 Chemical information provided by USEPA Chemicals D
Year	2023
Version	Final
Objective	Final IRIS Toxicological Review of Perfluorohexanoic Acid (PFHxA, CASRN information, visit EPA's IRIS website for PFHxA.
Authors	U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment
Vocabulary	EPA Environmental health vocabulary

Assessment endpoints (1082 found)

Filter by endpoint name (ex: heart weight) | ↑ study | 25 per page

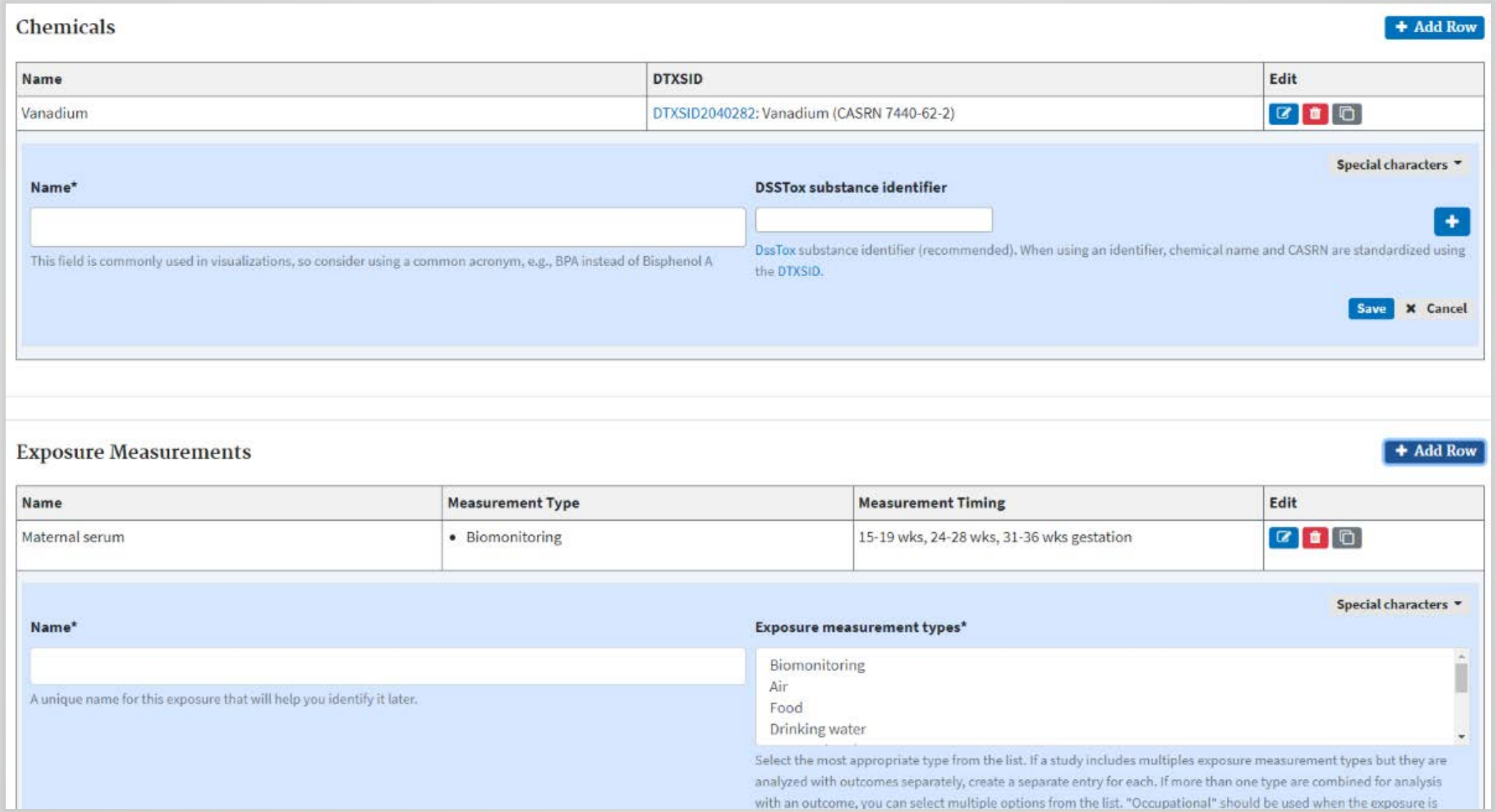
Study	Experiment	Animal group	Endpoint	Units	NEL	LEL	BMD	BMDL
Chengelis, 2009, 2850404	90-Day Oral	Male Crl:CD(SD) Rats	Peroxisomal Beta Oxidation	mg/kg-day	-	200.0	-	-
Chengelis, 2009, 2850404	90-Day Oral	Male Crl:CD(SD) Rats	Liver Weight, Relative to Brain	mg/kg-day	50.0	200.0	-	-
Chengelis, 2009, 2850404	90-Day Oral	Male Crl:CD(SD) Rats	Kidney Weight, Relative to Brain	mg/kg-day	200.0	-	-	-

- The HAWC page for the IRIS program's toxicological review of PFHxA
  - Includes extensive information such as individual study evaluations and visualizations for risk of bias and extracted data from relevant studies

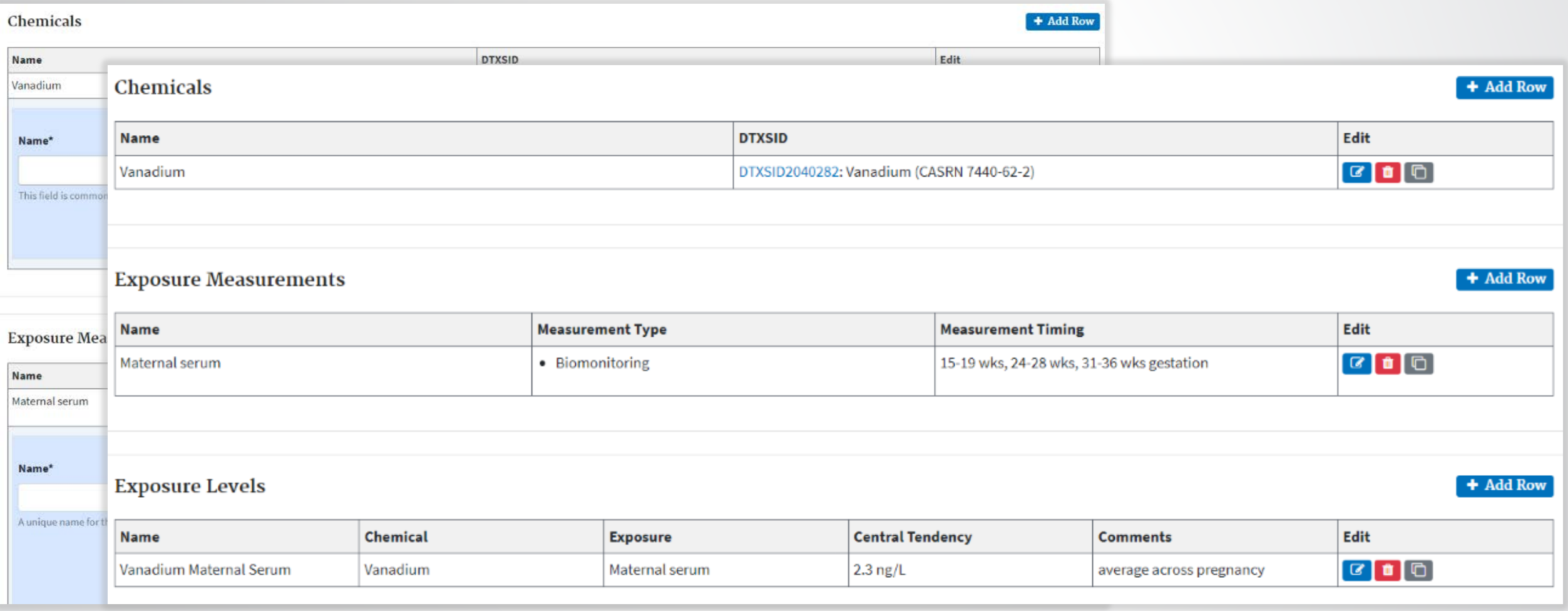
Interoperability with Other Tools

- Data can be moved from other systematic review tools into HAWC. Data can be accessed through API to support visualizations or data analysis

Epidemiology Data Extraction Module



Epidemiology Data Extraction Module



Chemicals

+ Add Row

Name	DTXSID	Edit
Vanadium	DTXSID2040282: Vanadium (CASRN 7440-62-2)	

Exposure Measurements

+ Add Row

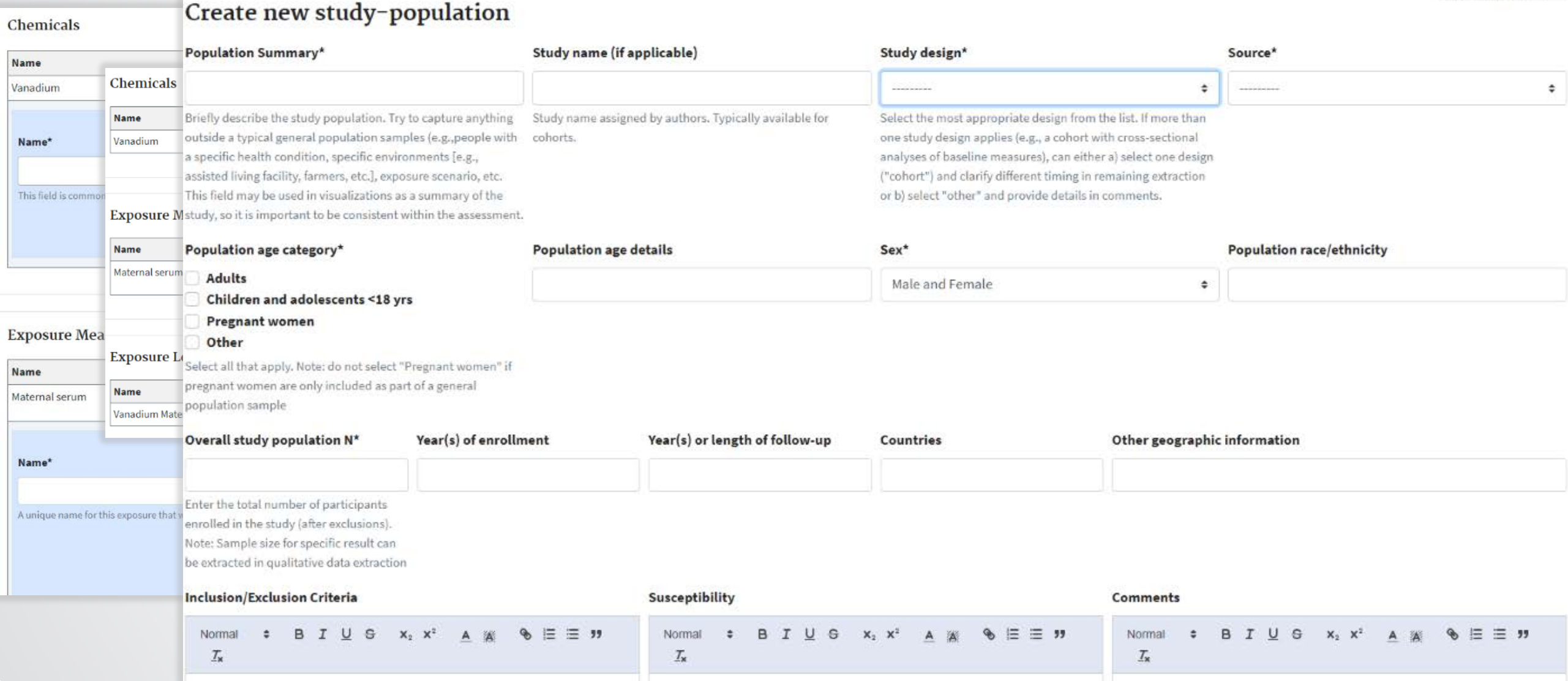
Name	Measurement Type	Measurement Timing	Edit
Maternal serum	• Biomonitoring	15-19 wks, 24-28 wks, 31-36 wks gestation	

Exposure Levels

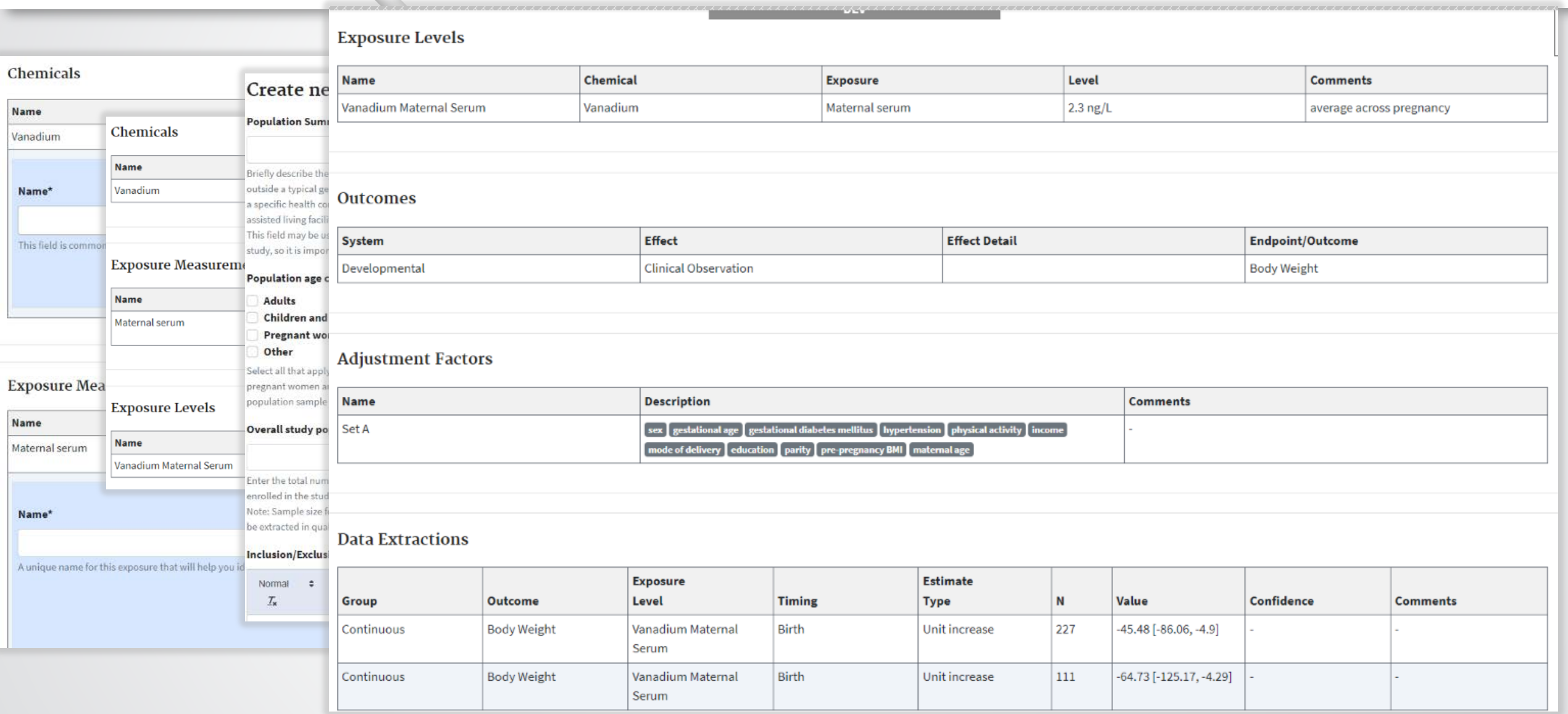
+ Add Row

Name	Chemical	Exposure	Central Tendency	Comments	Edit
Vanadium Maternal Serum	Vanadium	Maternal serum	2.3 ng/L	average across pregnancy	

Epidemiology Data Extraction Module



Epidemiology Data Extraction Module



Chemicals

Name

Vanadium

Name*

This field is common

Chemicals

Name

Vanadium

Exposure Measurem

Name

Maternal serum

Exposure Levels

Name

Vanadium Maternal Serum

Exposure Mea

Name

Maternal serum

Name*

A unique name for this exposure that will help you id

Create ne

Population Sumi

Briefly describe the
outside a typical ge
a specific health co
assisted living facili
This field may be us
study, so it is impor

Population age c

Adults

Children and

Pregnant wo

Other

Select all that apply
pregnant women a
population sample

Overall study po

Enter the total num
enrolled in the stud
Note: Sample size f
be extracted in qua

Inclusion/Exclus

Normal

Exposure Levels

Name	Chemical	Exposure	Level	Comments
Vanadium Maternal Serum	Vanadium	Maternal serum	2.3 ng/L	average across pregnancy

Outcomes

System	Effect	Effect Detail	Endpoint/Outcome
Developmental	Clinical Observation		Body Weight

Adjustment Factors

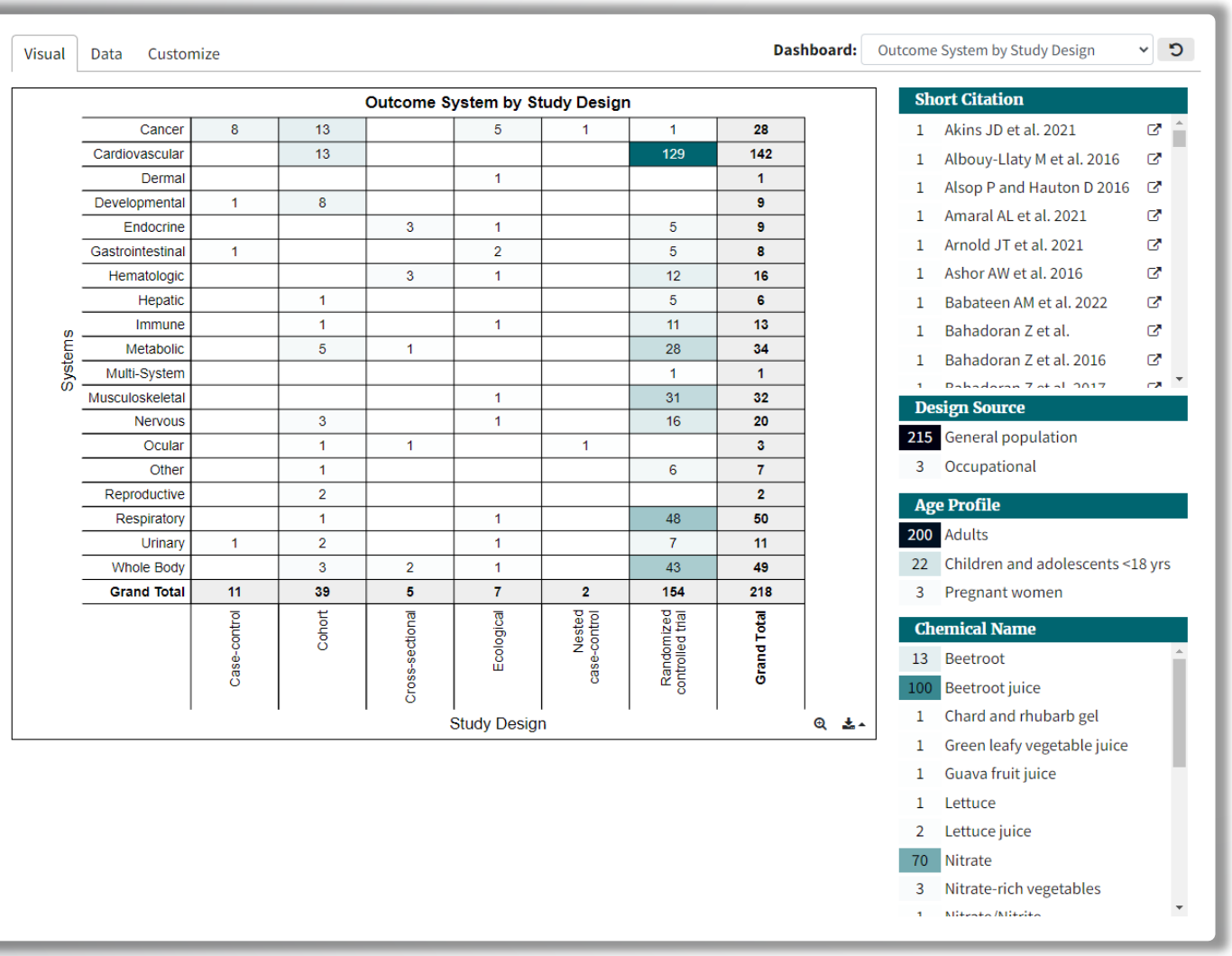
Name	Description	Comments
Set A	sex, gestational age, gestational diabetes mellitus, hypertension, physical activity, income, mode of delivery, education, parity, pre-pregnancy BMI, maternal age	-

Data Extractions

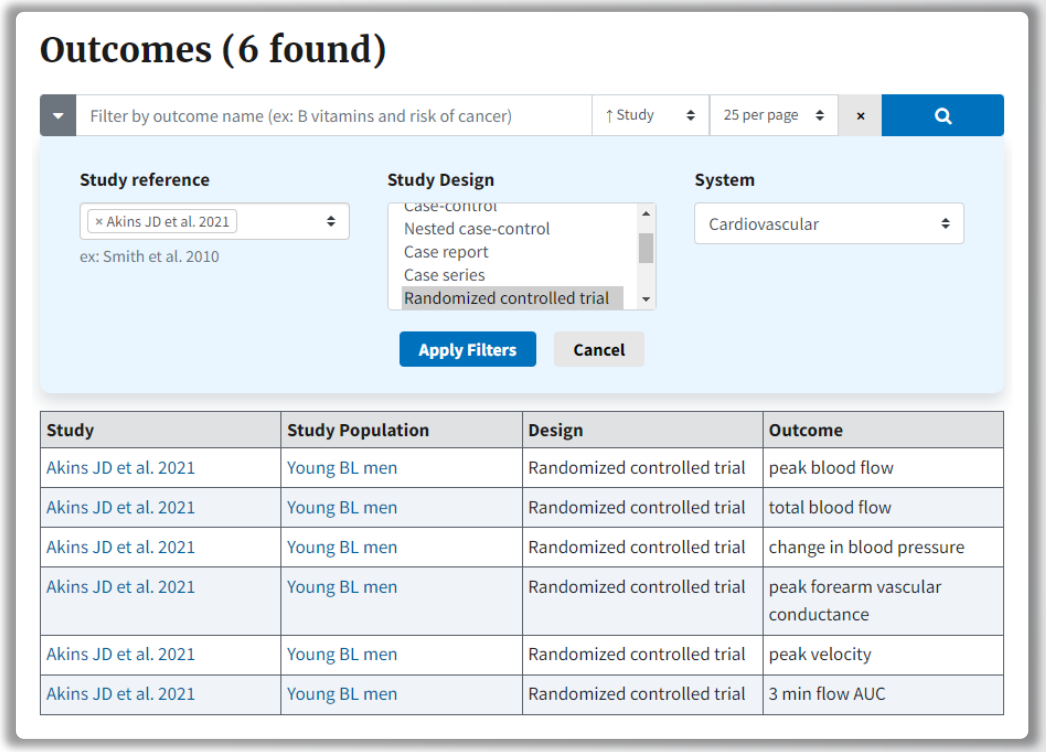
Group	Outcome	Exposure Level	Timing	Estimate Type	N	Value	Confidence	Comments
Continuous	Body Weight	Vanadium Maternal Serum	Birth	Unit increase	227	-45.48 [-86.06, -4.9]	-	-
Continuous	Body Weight	Vanadium Maternal Serum	Birth	Unit increase	111	-64.73 [-125.17, -4.29]	-	-

Epidemiology Data Extraction Module: Visualization Integration

Data visualizations

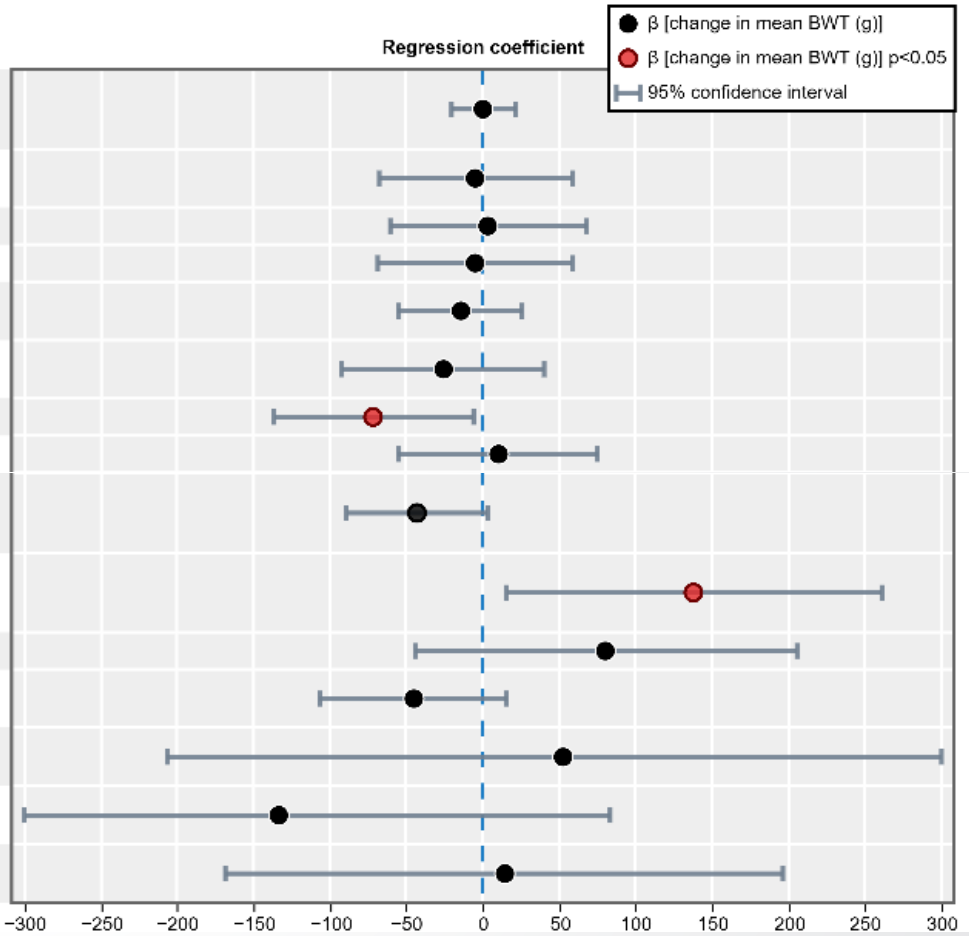


Data querying



Epidemiology Data Extraction Module: Visualization Integration

Study	Population	Overall Study Confidence	Design	Exposure Window	Regression Coefficient	Exposure Comparison
Buck Louis, 2018, 5016992	NICHD Fetal Growth Studies (2009-2013), United States, 2106 mother-infant pairs	|High|	Cohort (Prospective)	Trimester 1	0.13	1 SD increase
Manzano-Salgado, 2017a, 4238465	INMA cohort (2003-2008) 1202 mother-infant pairs	|High|	Cohort (Prospective)	Trimester 1	-4.75	Quartile 2
					3.56	Quartile 3
					-5.34	Quartile 4
					-14.82	ln-unit (ng/mL) increase
Bach, 2016, 3981534	Aarhus Birth Cohort (2008-2013), Denmark, 1507 mother-infant pairs	|High|	Cohort (Prospective)	Trimester 1-2	-26	Quartile 2
					-72	Quartile 3
					10	Quartile 4
Lenters, 2016, 5617416	INUENDO (2002-2004), Greenland/Poland/Ukraine, 1,321 mother-infant pairs	|Medium|	Cohort (Prospective)	Trimester 2-3	-43.45	ln-unit (ng/mL) increase
Cao, 2018, 5080197	Zhoukou City Longitudinal Birth Cohort (2013-2015), China, 282 mother-infant pairs	|Low|	Cohort (Prospective)	At birth	138.1	Tertile 2
					80.4	Tertile 3
Li, 2017, 3981358	GBCS (2013), China, 321 mother-infant pairs	|Low|	Cross-sectional	At birth	-45.6	ln-unit (ng/mL) increase
Shi, 2017, 3827535	Haidan Hospital (2012) 170 mother-infant pairs	|Low|	Cross-sectional	At birth	52.68	log10-unit (ng/mL) increase
Xu, 2019, 5381338	Cross-sectional study (2016-2017), China, 98 mother-infant pairs	|Low|	Cross-sectional	At birth	-133.4	ln-unit (ng/mL) increase
Callan, 2016, 3858524	AMETS (2008-2011), Australia, 98 mother-infant pairs	|Low|	Cross-sectional	Trimester 3	14	ln-unit (ng/mL) increase

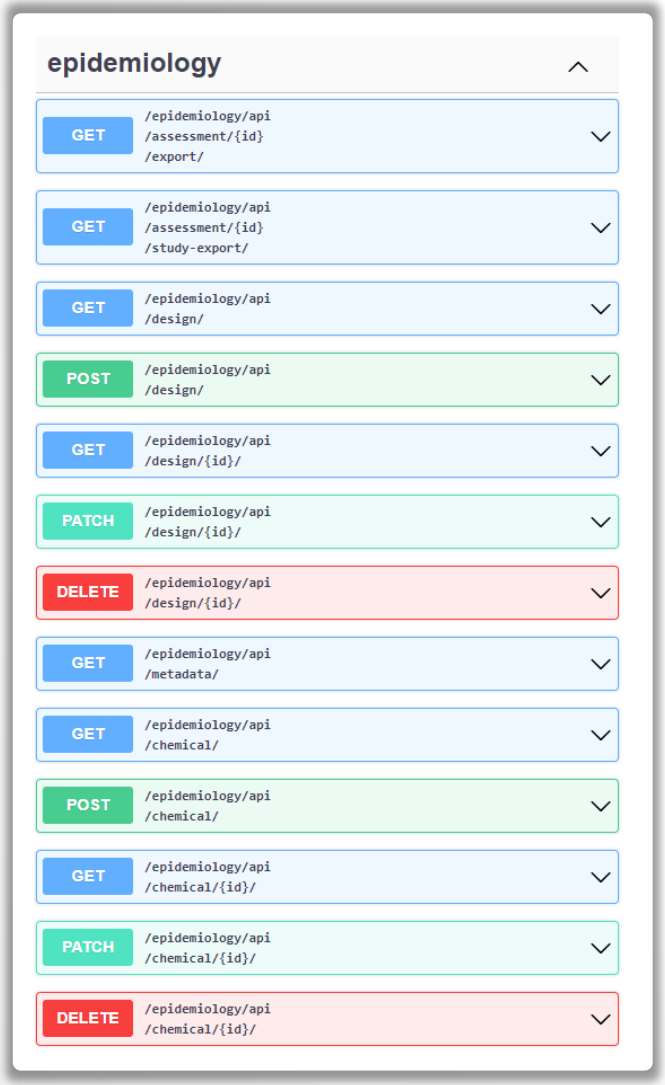


Data visualization: Example of a forest plot showing results from multiple studies evaluating changes in birth weight

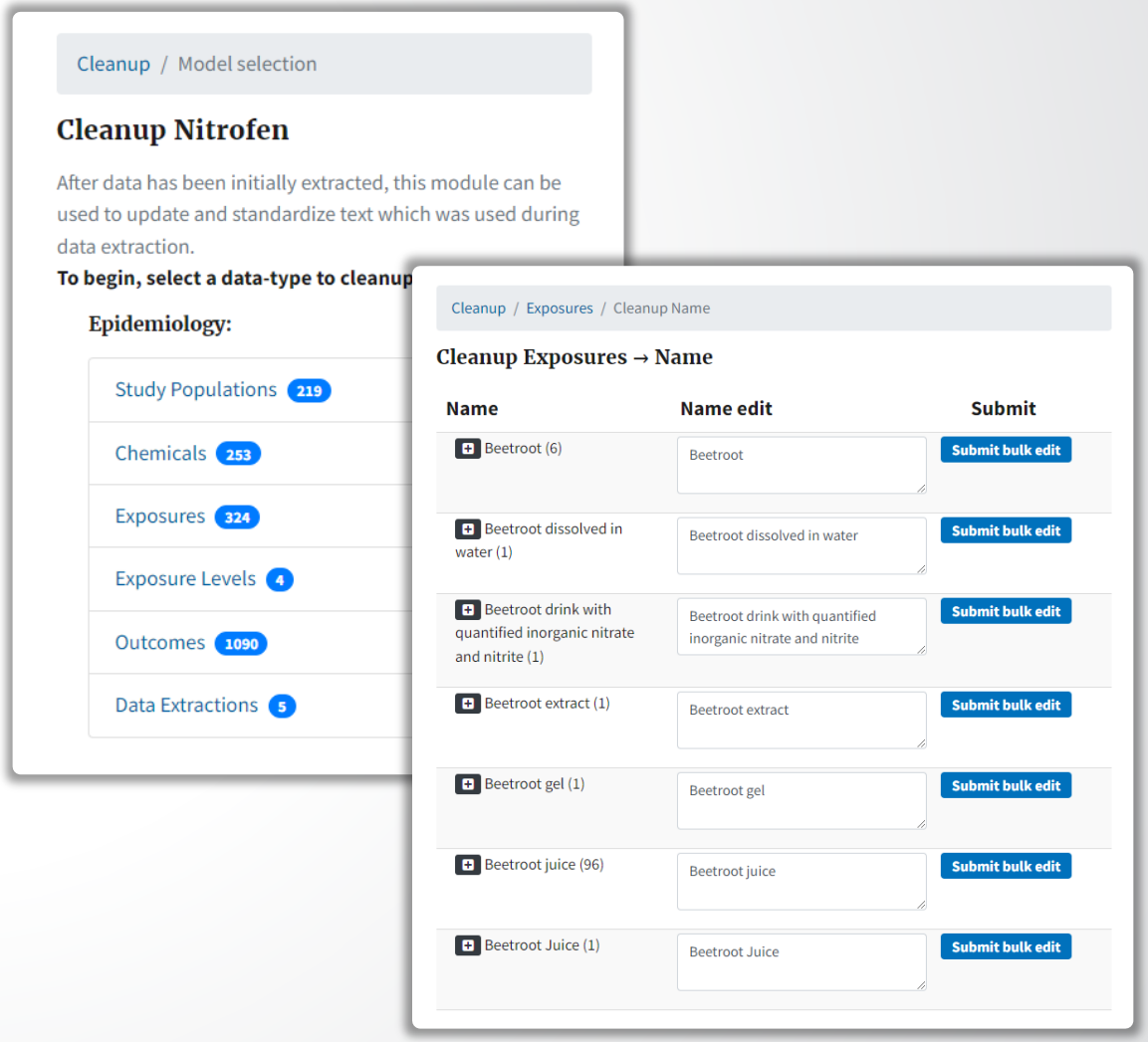
Epidemiology Data Extraction Module: API and advanced cleanup

- Added an application programming interface (API) layer to be able to create all objects via the API
- Added bulk-cleanup features to cleanup manually extracted data and standardize

API



Bulk cleanup



Interactive Visuals

Population	Outcome	Comments	Exposure Contrast (EWPM)	adjOR
USA (TX 1996-2[illegible] case-control, ♀♂ (60,613 case-mothers; 244,927 control-mothers)	anencephaly	3,985 (1.6%) controls and 10 (1.7%) cases[illegible]0 group	0 vs >0	1.09

Participant Selection	Exposure Measurement	Outcome Ascertainment	Confounding	Analysis	Selective Reporting
+	-	++	+	-	+

Click on any cell above to view details.

✚ Show all details

Data type(s)	Epidemiology
Full citation	Maternal residential proximity to chlorinated solvent emissions a[illegible] study. Brender JD, Shinde MU, Zhan FB, Gong X, Langlois PH. Env
Abstract	**BACKGROUND:** Some studies have noted an association betwee[illegible] chlorinated solvents and birth defects in offspring, but data are l[illegible] air emissions of these solvents on birth defects. **METHODS:** With data from the Texas Birth Defects Registry for bi[illegible] the relation between maternal residential proximity to industrial[illegible] birth defects in offspring of 60,613 case-mothers and 244,927 con[illegible] exposures to solvent emissions were estimated with metrics that[illegible] industrial sources and annual amounts of chemicals released. Lo[illegible] ratios and 95% confidence intervals for the associations between[illegible] chlorinated solvents and selected birth defects, including neural[illegible] congenital heart defects. All risk estimates were adjusted for year[illegible]

spina bifida

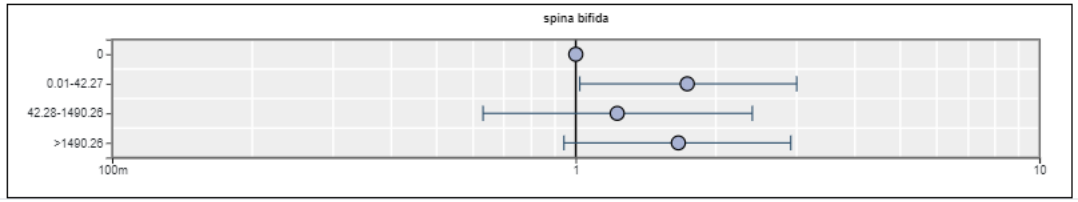
Population description	p-t[illegible]027; 9-14 (0.7-1.1%) exposed cases per group
Metric Description	adjOR
Adjustment factors	birth year geographic maternal age race/ethnicity
Dose response	not-applicable
Statistical power	not reported or calculated
Prevalence incidence	Number of cases: 1276 (97.2%) for referent, 14 (1.1%) for 0.001-42.27, 9 (0.7%) for 42.28-1490.26, 14 (1.1%) for >1490.26

Results by group

Group[a]	N	Estimate (---)	95% confidence intervals	*p*-value
0	239716	1	-	
0.01-42.27	1345	1.74	1.02 – 2.99	
42.28-1490.26	1340	1.23	0.63 – 2.4	
>1490.26	1337	1.66	0.94 – 2.91	

[a] Trend-test result: 0.027.

Forest plot



Summary

- IRIS assessments are developed using systematic review methods, which can be time and resource intensive
- Software tools such as HAWC can aid in streamlining multiple steps of assessment development, including the extraction of data from epidemiology studies. These data extractions are integrated with visualization capabilities and can be produced with minimal data processing.
- The flexibility provided with the updated data extraction in HAWC enables the development and storage of information for evidence maps, toxicological reviews, visualizations, and interoperability with other tools.

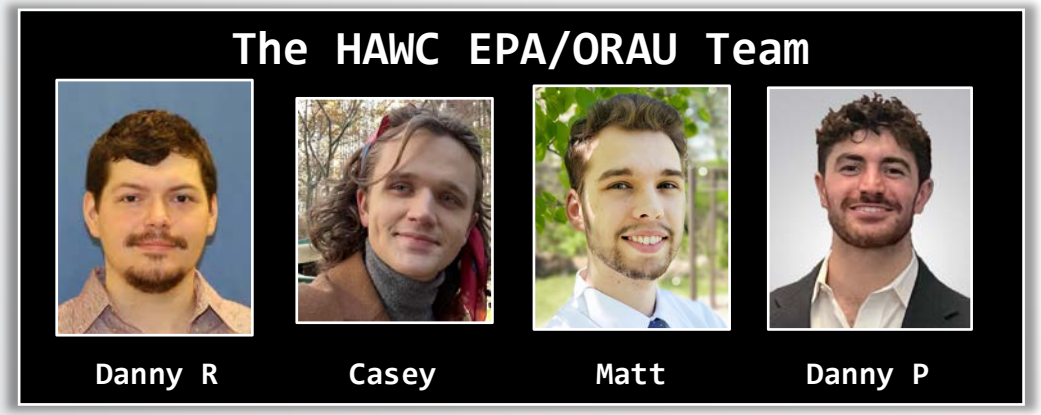
Acknowledgements

Code & Product Contributors

Casey Hansen (EPA/ORAU)
Matt Munns (EPA/ORAU)
Danny Peterson (EPA/ORAU)
Danny Rabstejnek (EPA/ORAU)
Carrie Smithing (EPA/ORAU)
Andy Shapiro (EPA)
Sean Watford (EPA)

Infrastructure & Support

Christine Alvarez (EPA)
Jaspreet Gill (GDIT)
Maureen Johnson (EPA)
Dahnish Shams (EPA)
Emily Snyder (EPA)
Michele Taylor (EPA)
Shane Thacker (EPA)
Kris Thayer (EPA)

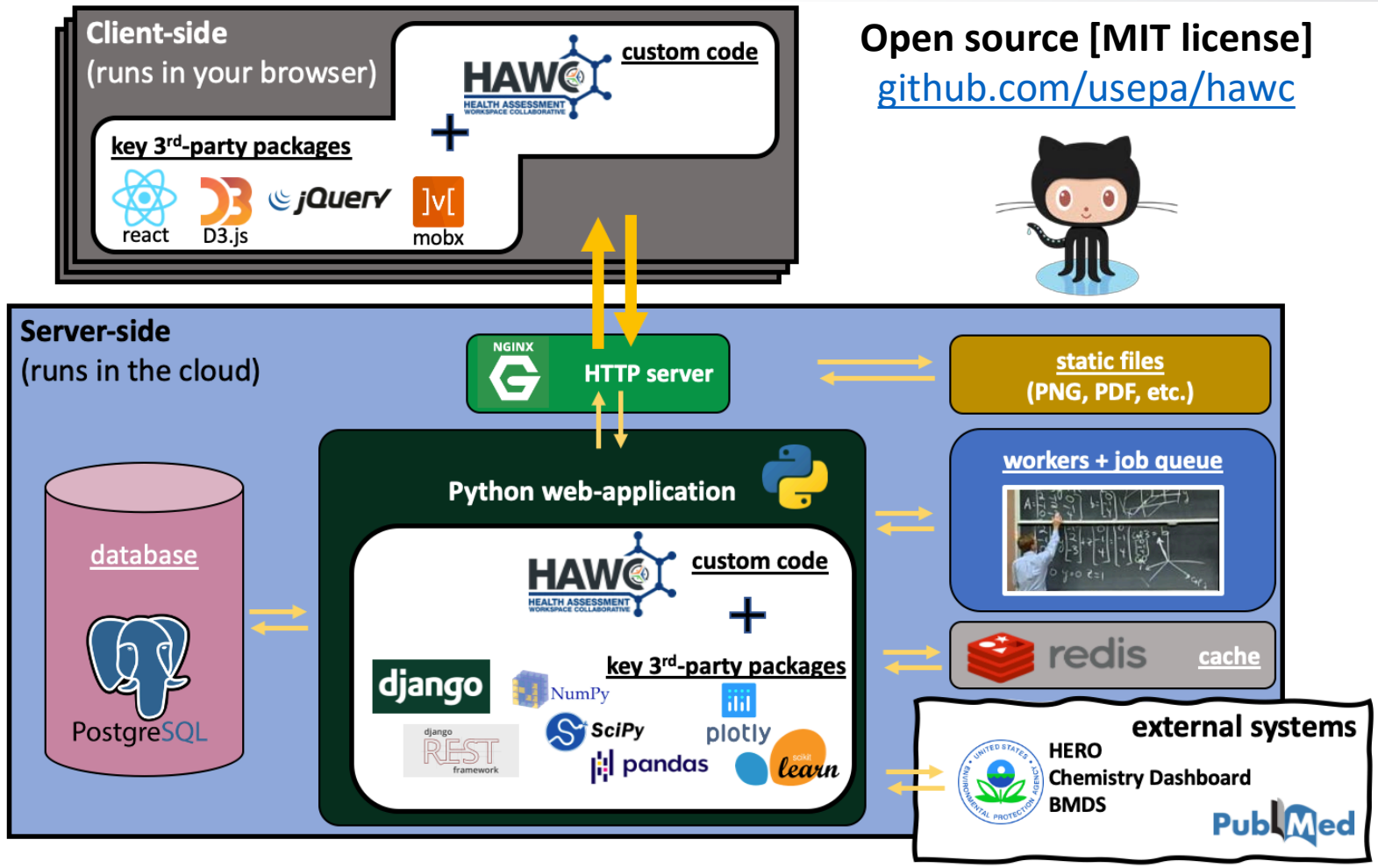


Extra Slides

So what is HAWC again?

Parlez-vous code?

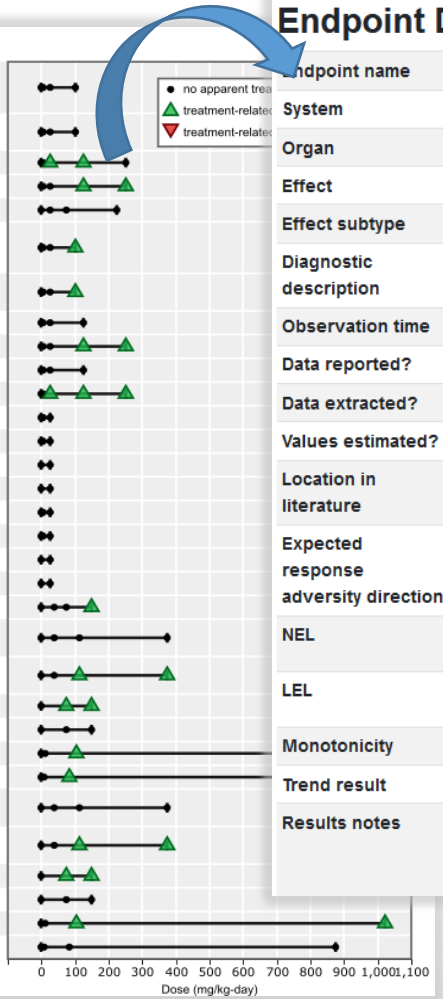
- A Python web-application
  - A web-application data entry in/out
  - APIs for automation of data in/out
  - Data science compute environment
- A relational database
  - Mostly relational data
  - Also binary/nosql data
- An interactive frontend
  - Dynamic visualizations
- An open-source application
  - Can collaborate with anyone



EPA Research and Development

Interactive Displays: Data Extraction

Chemical	Endpoint	Study	Animal Description	Route	Exposure Duration
6:2 Fluorotelomer alcohol	Liver Weight, Absolute	Mukerji et al. 2015	P0 Mouse, Crl:CD-1(ICR)BR (♀)	oral gavage	14d pre-mating, 14d mating, gestation, lactation
			P0 Mouse, Crl:CD-1(ICR)BR (♂)	oral gavage	109 d (premating-sacrifice)
		Serex T et al. 2014	Rat, Crl:CD(SD) (♀)	oral gavage	90 d
			Rat, Crl:CD(SD) (♂)	oral gavage	90 d
		Unnamed report (2005a) (ECHA summary)	Rat, Crl:CD(SD) (♂♀)	oral gavage	28 d
	Liver Weight, Relative	Mukerji et al. 2015	P0 Mouse, Crl:CD-1(ICR)BR (♀)	oral gavage	14d pre-mating, 14d mating, gestation, lactation
			P0 Mouse, Crl:CD-1(ICR)BR (♂)	oral gavage	109 d (premating-sacrifice)
		ECHA, 2007, 5701160	Rat, Crl:CD(SD) (♀)	oral gavage	28d (1dose/d)
		Serex T et al. 2014	Rat, Crl:CD(SD) (♀)	oral gavage	90 d
		ECHA, 2007, 5701160	Rat, Crl:CD(SD) (♂)	oral gavage	28d (1dose/d)
		Serex T et al. 2014	Rat, Crl:CD(SD) (♂)	oral gavage	90 d
6:2 Fluorotelomer methacrylate	Liver Weight, Absolute	ECHA, 2007, 6299223	Rat, Crl:CD(SD) (♀)	oral gavage	28d (1dose/d)
			Rat, Crl:CD(SD) (♂)	oral gavage	28d (1dose/d)
	Liver Weight, Absolute, Recovery	ECHA, 2007, 6299223	Rat, Crl:CD(SD) (♀)	oral gavage	28d (1dose/d)
			Rat, Crl:CD(SD) (♂)	oral gavage	28d (1dose/d)
	Liver Weight, Relative	ECHA, 2007, 6299223	Rat, Crl:CD(SD) (♀)	oral gavage	28d (1dose/d)
			Rat, Crl:CD(SD) (♂)	oral gavage	28d (1dose/d)
	Liver Weight, Relative, Recovery	ECHA, 2007, 6299223	Rat, Crl:CD(SD) (♀)	oral gavage	28d (1dose/d)
			Rat, Crl:CD(SD) (♂)	oral gavage	28d (1dose/d)
Trifluoroacetic acid	Liver Weight, Absolute	Unnamed Report (2010a) (ECHA Summary)	P0 Rat, Crl:CD(SD)IGS BR (♀)	oral gavage	GD 6-19
		Unnamed Report (2012b) (ECHA Summary)	P0 Rat, Crl:CD(SD)IGS BR (♀)	oral gavage	up to 57 d (premating-lactation)
			P0 Rat, Crl:CD(SD)IGS BR (♂)	oral gavage	38 d (premating-termination)
		Saillenfait et al. 1997	P0 Rat, Sprague-Dawley (♀)	oral gavage	GD 10-20
			F1 Rat, Sprague–Dawley (♂♀)	oral gavage	GD 10-20
		Unnamed Report (2016a) (ECHA Summary)	Rat, Wistar Rj:Wi (Iops Han) (♀)	oral diet	90 d
			Rat, Wistar Rj:Wi (Iops Han) (♂)	oral diet	90 d
	Liver Weight, Relative	Unnamed Report (2012b) (ECHA Summary)	P0 Rat, Crl:CD(SD)IGS BR (♀)	oral gavage	up to 57 d (premating-lactation)
			P0 Rat, Crl:CD(SD)IGS BR (♂)	oral gavage	38 d (premating-termination)
		Saillenfait et al. 1997	P0 Rat, Sprague-Dawley (♀)	oral gavage	GD 10-20
			F1 Rat, Sprague–Dawley (♂♀)	oral gavage	GD 10-20
		Unnamed Report (2016a) (ECHA Summary)	Rat, Wistar Rj:Wi (Iops Han) (♀)	oral diet	90 d
			Rat, Wistar Rj:Wi (Iops Han) (♂)	oral diet	90 d



Liver Weight, Absolute

Endpoint Details

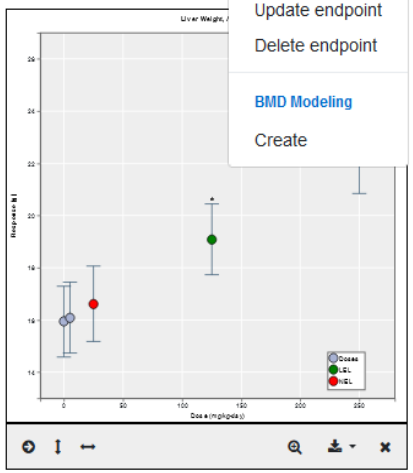
Endpoint name	Liver Weight, Absolute
System	Hepatic
Organ	Liver
Effect	Clinical Observation
Effect subtype	Organ Weight
Diagnostic description	Liver, Weight
Observation time	90 d
Data reported?	✔
Data extracted?	✔
Values estimated?	–
Location in literature	Table 5
Expected response adversity direction	---
NEL	25 mg/kg-day
LEL	125 mg/kg-day
Monotonicity	--
Trend result	not reported
Results notes	"Following 90 days of dosing, effects on organ weights were present in the testes, liver and kidney of males (Table 5) and in livers and kidneys

Plot

Actions

Endpoint Editing
Update endpoint
Delete endpoint

BMD Modeling
Create



Dataset

Dose (mg/kg-day)	Number of Animals	Response (g)	Standard Deviation
0	10	15.94	1.9
5	10	16.09	1.9
25[a]	10	16.62	2.02
125[b,c]	10	19.09	1.89
250[b]	8	22.84	2.39

[a] NEL (No effect level)
[b] Significantly different from control ($p < 0.01$)
[c] LEL (Lowest effect level)