medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

The GeoTox Package: Open-source software for connecting spatiotemporal exposure to individual and population-level risk

Kyle P Messier[1,2*], David M Reif[1] and Skylar W Marvel[1]

[1*]Predictive Toxicology Branch, Division of Translational Toxicology, National Institute of Environmental Health Sciences, 530 Davis Dr, Durham, 27713, NC, USA.
[2]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, 111 T.W. Alexander Dr, Research Triangle Park, 27709, NC, USA.

*Corresponding author(s). E-mail(s): kyle.messier@nih.gov;
Contributing authors: david.reif@nih.gov; skylar.marvel@nih.gov;

Abstract

Background: Comprehensive environmental risk characterization, encompassing physical, chemical, social, ecological, and lifestyle stressors, necessitates innovative approaches to handle the escalating complexity. This is especially true when considering individual and population-level diversity, where the myriad combinations of real-world exposures magnify the combinatoric challenges. The GeoTox framework offers a tractable solution by integrating geospatial exposure data from source-to-outcome in a series of modular, interconnected steps.
Results: Here, we introduce the **GeoTox** open-source R software package for characterizing the risk of perturbing molecular targets involved in adverse human health outcomes based on exposure to spatially-referenced stressor mixtures. We demonstrate its usage in building computational workflows that incorporate individual and population-level diversity. Our results demonstrate the applicability of GeoTox for individual and population-level risk assessment, highlighting its capacity to capture the complex interplay of environmental stressors on human health.
Conclusions: The **GeoTox** package represents a significant advancement in environmental risk characterization, providing modular software to facilitate the application and further development of the GeoTox framework for quantifying the relationship between environmental exposures and health outcomes.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

By integrating geospatial methods with cutting-edge exposure and toxicological frameworks, **GeoTox** offers a robust tool for assessing individual and population-level risks from environmental stressors. **GeoTox** is freely available at https://niehs.github.io/GeoTox/.

Keywords: exposome, risk, population susceptibility, source-to-outcome

1 Introduction

Risk characterization of multiple environmental stressors including physical, chemical, social, ecological, and lifestyle, is a top priority of the environmental health sciences [1]. There are numerous technical factors contributing to the complexity of the problem, including the infinite number of real-world exposure combinations and the inter-individual biological complexity of humans. At the intersection of exposure mixtures and biological complexity is the well-established result that most human diseases occur due to high-dimensional interactions between exogenous environmental exposures and endogenous biology such as genetics, epigenetics, and the microbiome [2].

To address the challenge of complex chemical exposures the *exposome* concept was introduced [3]. The exposome (exposomics) is measure of all environmental, social, lifestyle, and ecological exposures across the life course of an individual [2]. The main-stream philosophy for quantifying the exposome is through individual biological measurements [4] such as biomarker concentrations, epigenetic alterations, metabalomic derivations, and proteomic responses [5]. Advancements in biological and -omic based quantification are happening quickly [6]; nonetheless, there are immense analytical challenges ahead in large part due to (1) the spatiotemporal variability of exposures, (2) the sheer magnitude and variety of exposures, (3) the specificity of exposomic measurements to multiple endogenous and exogenous processes, and (4) the temporal variability and specificity of biological responses.

Geospatial approaches offer an attractive approach to quantify the exposome because external components such as the social and physical-chemical exposome are more accurately quantified with geospatial models of exposure. Described by Rappaport and Smith [7] and supported by Vermeulen et al. [2], toxic effects are mediated through chemicals that alter critical molecules, cells, and physiological processes inside the body. In the same spirit, it is clear that phenotypic outcomes (e.g. disease) at the individual or population level can only occur after a series of source, exposure, and biological dynamics (Figure 1). We refer to this as the *source-to-outcome-continuum*. It follows that if methods and data exist to quantify each step in the sequence, and each step can be integrated into each neighboring step, then individual and population outcomes can be quantified from spatiotemporally resolved environmental sources and exposures.

Conveniently, data and methods do exist for quantifying the source-to-outcome-continuum. The Adverse Outcome Pathway (AOP) [8] framework provides a linkage

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

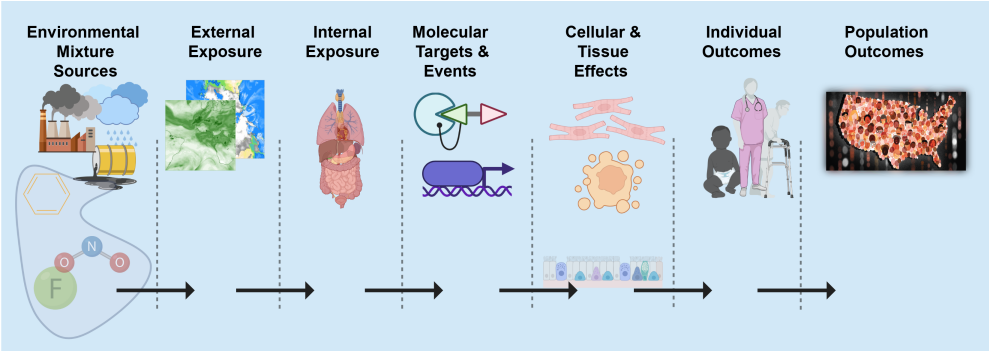

Fig. 1: **Source-to-outcome-continuum**: A schematic showing the necessary sequence of events such that each step must occur for environmental exposures to cause individual and population outcomes

between perturbation of a specific biological target, pathway or process by a stressor, and an adverse outcome considered relevant to risk assessment. Many sources of data can be used to support AOP development, with new approach methodologies (NAMs) such as high-throughput in-vitro assays, small organism models, and in silico toxicology modeling approaches especially useful for probing specific targets. Teeguarden et al. [9] proposed the Aggregate Exposure Pathway (AEP) as an exposure analog to the AOP framework. The AEP framework aims to quantify the fate and transport of environmental chemicals through different media and, when appropriate, chemical transformations that occur in the environment. Additionally, the framework articulates the adsorption, distribution, metabolism, and excretion of chemicals that relate the external concentration and internal concentration of a chemical (or its active metabolites). Most importantly, the AEP framework defines these internal concentrations as the Target Site Exposures (TSEs), which are analogous to molecular initiating event (MIE) from the AOP framework. To summarize, the end process in an AEP is the beginning process for the AOP thus providing explicit methods for source-to-outcome-continuum modelling.

Recent work has demonstrated the potential for *source-to-outcome-continuum* modeling in NAMs based risk assessment. For example, Hines et al. [10] estimated a hazard index for humans, fish, and small mammals at hypothetical field site with exposures to ClO_4^- and its degradation products. Price et al. [11] posit *source-to-outcome-continuum* modeling as a more comprehensive approach for integrating chemical interactions in human health risk assessment. In Eccles et al. [12], we introduced a workflow, referred to as GeoTox, for characterizing the risk of perturbing molecular targets involved in adverse human health outcomes based on exposure to spatially referenced chemical mixtures by connecting the frameworks of the AEP and AOP. The objectives of this paper build upon the GeoTox framework as follows: (1) Develop open-source software to facilitate analysis of the geospatial toxicological workflow that features best-practices in software development including continuous integration, unit testing, ready tunability of key parameters, documentation, and (2)

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Demonstrate with case-studies of both individual and populations, the risk as quantified by single and multiple assays end-points in the source-to-outcome framework. Our goal is that the **GeoTox** software will allow researchers in the toxicological and geospatial research communities to build and expand application of the GeoTox concept.

The remainder of this manuscript is as follows: In section 2 we provide a review of the GeoTox framework and an overview of the open-source **GeoTox** package to facilitate usage across the environmental and toxicological risk assessment community (Section 2.1). Section 2.2 provides details on the **GeoTox** package functions. Section 2.3 introduces the population and individual level analysis and section 3 summarizes the case study results. Section 4 provides a discussion on the **GeoTox** code development and case study results. Lastly, section 5 summarizes the package features and places it in the context of new approach methodologies development for improved exposomic risk characterization.

2 Materials and Methods

2.1 Overview of GeoTox framework and package

Here, we provide a brief review of the GeoTox framework as described in Eccles et al. [12] including its key steps and data requirements. Additionally, we describe the key functionality in the **GeoTox** package. Throughout the manuscript, software names and functions are highlighted in **code text.** This differentiates **GeoTox** the software from GeoTox the framework.

Figure 2 is an infographic describing the key steps in the workflow and summaries of the package functionality. First, geospatial modeling is used to estimate an external, space-time referenced exposure concentrations due to presumed external sources. Eccles et al. [12] utilized publicly available chemical transport model external exposure data, but all external exposure assessment methods (e.g. geostatistics, chemical transport, personal sensors), media (e.g. outdoor air, indoor air, groundwater), and routes of exposure (e.g. inhalation, oral ingestion) are amenable. Importantly, **GeoTox** is not a geospatial exposure assessment tool as that is out of scope, thus it is necessary to have georeferenced exposure data for the chemicals of interest. Nonetheless, a primary purpose of the package maintains is to provide explicit geospatial data (e.g. point locations, census boundaries) linkage with proceeding analysis steps. To ensure downstream analysis joins, the external geospatial exposure chemical identifiers (e.g. CASRN, SMILES) must also be available in the subsequent databases and model. For example, while ozone (O_3) is an established air pollutant with readily available geospatial exposure data, it is currently not amenable to PBPK models or high-throughput in vitro screening assays. Likewise, many novel or bespoke chemicals have in vitro screening data but are not monitored or modeled for geospatial exposures thus a join is not currently possible.

Next, behavioral and physiological modeling is used to estimate internal concentrations from the presumed route of exposure. The standard source is the EPA Exposure Factors Handbook [13], which provides population distributions of physiological factors

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

such as inhalation rates by age, sex, and body weight. The physiological parameters are also provided as population distributions by census areal units from the US Census Bureau, thus providing the explicity geospatial connection from external to internal exposure.

Next, physiologically-based toxicokinetic (PBTK or simply TK) models are used to estimate the internal, target organ concentration. The R package, **httk** [14], is the standard used in **GeoTox** due to its high-throughput nature and flexibility, but tailored PBTK models are also applicable. Then, in-vitro-to-in-vivo-equivalent (IVIVE) modeling, also based on PBTK models, is used to convert the human concentration to an in-vitro equivalent concentration.

With the external exposure data now comparable to the high-throughput screening (HTS) assay data, individual and mixtures concentration-response modeling can be performed. The HTS assay or assays chosen inform on the molecular target of interest through a molecular initiating event and is both biologically and risk relevant (e.g. Related to a cancer or outcome mode of action). The molecular or biological target such as a molecular initating event or adverse outcome pathway key event anchors the framework as a mechanistic-based risk mapping approach. While traditional in-vivo data can be used, the examples here draw from the Integrated Chemical Environment (ICE) curated high-throughput in-vitro screening (cHTS) [15]. Section 2.2.8 discusses the availability and integration of cHTS data in **GeoTox** and section 4 discusses other data sources and potential future integration. The **GeoTox** package has functionality for maximum likelihood fitting of 2 and 3 parameter hill dose-response models for single chemical and assay combinations. The total mixture response is also estimated using either generalized concentration addition (GCA) or independent action (IA).

Lastly, a risk assessment metric is chosen for the risk assessment, resulting in the final risk map. Eccles et al. [12] showed risk maps of the mixture response predicted by independent action (IA) [16] and generalized concentration addition (GCA) [17] and the hazard quotient (HQ)[18], which are all available risk metrics in **GeoTox**. Total mixture response calculated from GCA or IA is useful for relative comparisons between geospatial locations or with reported values from previous studies. The HQ is a unitless ratio that compares the GeoTox based predicted response to some reference response value, which can be useful for chemical prioritization or risk-based decision making.

2.2 GeoTox Package Details

2.2.1 FAIR Code Development

GeoTox is created with many software best-practices, including findable, accessible, interoperable, and reproducible (FAIR) software development standards. The package is accessible as the source code is made completely open-source, hosted at the NIEHS GitHub (code base: https://github.com/NIEHS/GeoTox; package website: https://niehs.github.io/GeoTox/). The NIEHS GitHub repository and availability through the Comprehensive R Archive Network (CRAN) ensure its potential for long-term findability and accessibility.

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

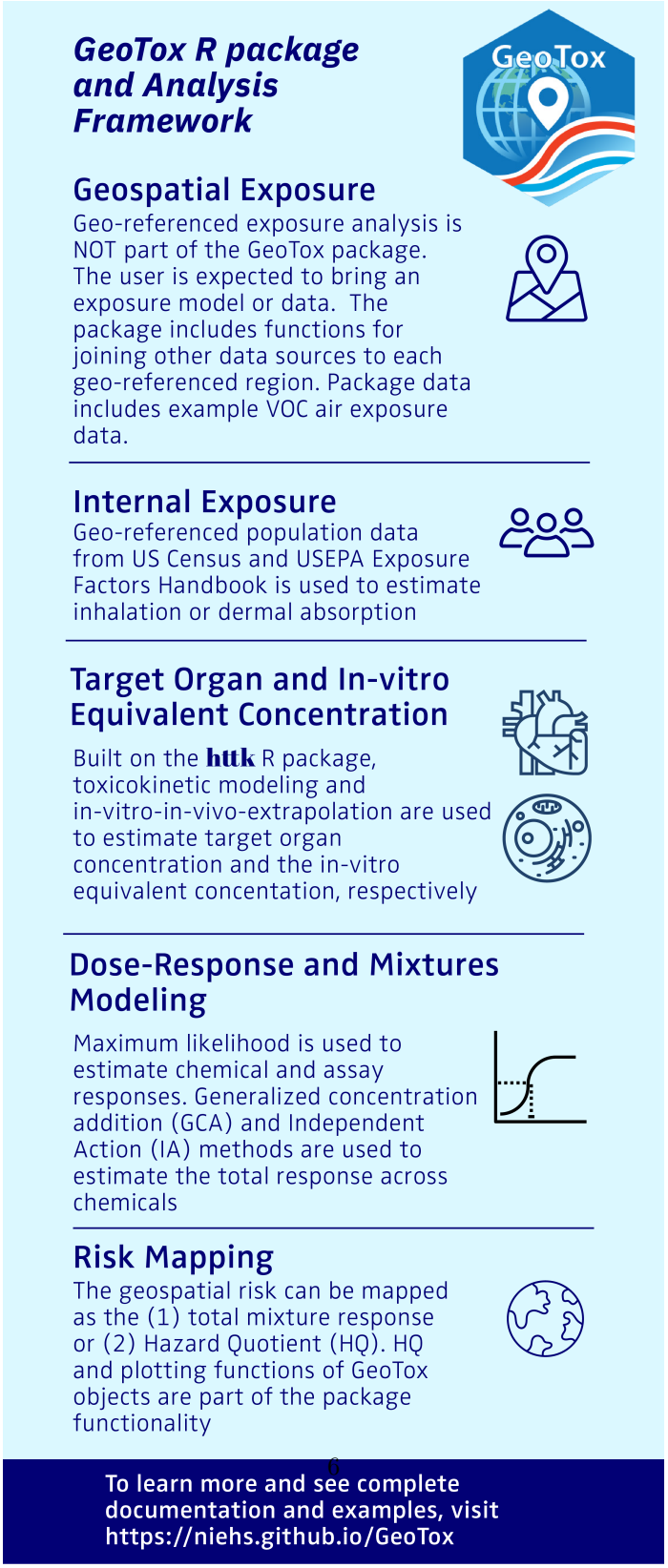

Fig. 2: GeoTox package and analysis framework infographic. The methods and package functionality are summarized for each analysis step.

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

GeoTox uses standardized style and linting through the **lintr** package [19] providing syntax consistency. The object-oriented approach is compatible with the **tidyverse** [20] syntax and style, allowing interoperability with R-based piping workflows (i.e. $x \mid > f()$). Tidy principles are a human-readable, consistent, composable, and inclusive syntax for data science workflows such as data import, exploratory analysis, cleaning, visualization, and model fitting [20].

GeoTox was intentionally developed in a modular and extensible manner allowing the authors and community members such as geospatial scientists and toxicologists to contribute and build upon the code base. As an open-source project under the MIT license, we provide a contributor's guide, news file for updates, and an on-going list of features or extensions that could be added by the authors or community members. An important contributing guide rule is the enforcement of continuous integration and continuous development (CI/CD) workflows on the GitHub repository. These workflows enforce rules and automatic checks to the code base when the authors or community members provide updates or new features. Additionally, we have created standardized templates for bug reports and feature requests on the GitHub repository. Together, these will aide in the on-going maintenance and further improvements while minimizing the overall technical debt.

Lastly, the **testthat** package is used to enforce test-driven development (TDD) in the creation and maintenance of **GeoTox** [21]. The unit tests verify that each function performs as designed and help minimize the chances of bugs - particularly semantic errors in which the program functions, but the calculation is not correct. Performance differs between functions and data sources, but can be generally categorized as expected successes, testing that the function does not return an error when provided valid parameters, and expected failures, testing that the function recognizes invalid parameters and returns an error.

2.2.2 GeoTox Object

The core development is the GeoTox object, a R-based S3 class object. S3 is R's simplest and most flexible version of object-oriented programming [22]. It's simplicity allows us to define the GeoTox object with flexible components that define each step in a GeoTox framework analysis. **sf** spatial objects, a data frame or tibble with a geometry list-column, provide georeferenced information to the GeoTox object and generic data frames and lists store information on each step in the GeoTox analysis framework.

Figure 3 shows an example GeoTox object printed to the R console. The header displays the geospatial regions and the simulated population within each region. Next, we see the object components listed as fields with their name, class, and dimensions. Details of each field are described in the following sections including the individual simulations based on the **httk** Monte Carlo population simulations from Ring et al. [23].

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

```
 GeoTox object
Assays: 229
Chemicals: 32
Regions: 100
Population: 500

 Data Fields:

Name            Class                       Dim
age             list(integer)               100 x (500)
IR              list(numeric)               100 x (500)
obesity         list(character)             100 x (500)
C_ext           list(matrix)                100 x (500 x 32)
C_ss            list(matrix)                100 x (500 x 32)

 Computed Fields:

Name            Class                       Dim
D_int           list(matrix)                100 x (500 x 32)
C_invitro       list(matrix)                100 x (500 x 32)
resp            list(data.frame)            100 x (114500 x 6)
sensitivity     list(list)                  5 x (100)

 Other Fields: par, boundaries, exposure, css_sensitivity, hill_params
```

Fig. 3: GeoTox object printed to console

2.2.3 Concentration-Response Model Fitting

There are many R packages available for toxicological concentration-response model fitting including **drc** [24], **tcpl2** [25], and **toxicR** [26, 27]. For internal consistency and completeness, we provide concentration-response model fitting for parametric hill models using maximum likelihood estimation comparable to **tcpl2**. The function *fit_hill()* takes a data frame consisting of concentration response data and fits a 2 or 3 parameter hill function. **GeoTox** package reference and vignettes describe how the data frame input is organized to comply with the function syntax. The output is stored in the GeoTox object as the *hill_params* field and is accessed when piped into the calculator functions (Section 2.2.5 for the subsequent analysis steps).

2.2.4 Simulate Functions

Monte Carlo sampling of known parameter distributions underlies the GeoTox framework to capture geospatially resolved individuals and populations. Age, obesity, inhalation rates, and steady-state blood plasma concentrations of chemicals are covered in the **GeoTox** simulation functions. Following the methods outlined in Eccles et al. [12], for a given region such as a US county, age distributions and obesity prevalence are obtained via the US Census and the Centers for Disease Control PLACES

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

data [28], respectively. Inhalation rates are simulated based on the age of the individuals and derived from Table 6.7 of the EPA Exposure Factors Handbook [13]. In section 4 we discuss potential future integration with more refined or user-defined inhalation rate data. Age and obesity simulation results are input into the toxicokinetic parameter simulation. The **httk** package is used to simulate toxicokinetic parameters such as C_{ss}, the steady-state plasma concentration in $\mu\dot{M}(\frac{mg}{kg})^{-1}$ [23]. Additionally, as the C_{ss} simulation can be time consuming for a large sample size, we simulate a tractable number of individuals using httk (i.e. $\leq$ 1,000) and then utilize bootstrap resampling for inference.

2.2.5 Calculator Functions

A series of self-described calculator functions are developed to perform the key steps of the GeoTox pipeline. Starting from the beginning of the pipeline, the *calc_internal_dose()* estimates the internal target organ dose of a chemical given inhalation rate, body weight, and time. Currently, time defaults to 1 since the estimate for blood plasma concentration is based on a steady-state assumption achieved in 1 day. *calc_invitro_concentration()* estimates the in vitro equivalent concentration via a simple multiplication of the internal dose and the steady-state blood plasma concentration. To calculate the total mixture response metrics used as the basis of the risk assessment, the function *calc_response()* estimates and returns the chemical mixture responses using the IA, GCA, and the HQ.

2.2.6 Sensitivity Analysis Functions

Every stage in the GeoTox framework includes uncertainty that contribute to the overall variability of the mapped mixture responses. The *compute_sensitivity()* functions wraps the Monte Carlo uncertainty analysis around each step of the analysis to calculate and aide in visualization of the overall uncertainty. By default, a GeoTox analysis function is chosen to vary freely while the remaining functions and their respective parameters are fixed to a central value. This simple sensitivity functionality in **GeoTox** will allow researchers to explore the sources and impacts of uncertainty in their analysis and increase confidence in the risk mapping results with targeted experiments.

2.2.7 Plotting Functions

Visualization is an important part of exploratory analysis and results dissemination in the GeoTox framework. Here, we provide easy and extensible plotting functions based on the GeoTox object. **GeoTox** provides a basic **plot()** function that takes the GeoTox object as input and can create **ggplot** based plots for geospatial exposure, **plot(geoTox, type = "exposure")**, dose-response results, **plot(geoTox, type = "hill")**, single and multi-assay responses, **plot(geoTox, type = "response")**, and sensitivity analysis results, **plot(geoTox, type = "sensitivity")**.

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

2.2.8 Package Data

The **GeoTox** package comes with a dataset, *geo_tox_data*, that is the basis for vignettes demonstrating the package functionality. Importantly, the "GeoTox Package Data" vignette shows the steps for downloading, processing, and creating the package dataset, which can easily be modified to make adjustments to underlying exposure and population data and will aide users in developing their own workflows based on novel datasets. The package dataset contains external, geospatial air pollution exposure data from the 2019 USEPA AirToxScreen, which utilizes the Community Multiscale Air Quality (CMAQ) and AERMOD dispersion models to estimate annual average hazardous air pollutants at the census tract level [29]. Chemical CAS numbers and preferred names are standardized and linked using metadata from the EPA CompTox dashboard [30]. Population age and obesity distributions are used from the US Census bureau and Center for Disease Control and Prevention PLACES data [28], respectively. Additionally, the code base includes example for downloading data sets curated for targeted toxicity endpoints by Integrated Chemical Environment (ICE) REST API [15, 31].

2.3 Case Studies

We showcase **GeoTox** with three novel analyses that demonstrate single assay, multi-assay, and individual risk characterization. The analyses follow the steps outlined in our package vignettes available at the package website.

First, we replicate the Eccles et al. [12] analysis in North Carolina, but utilize an H2AX histone modification genotoxicity assay (assay id: TOX21_H2AX_HTRF_CHO_Agonist_ratio) which is a demonstrated in-vitro marker for genotoxic effects and potential in-vitro replacement for in-vivo based cancer relative potency factors [32]. This analysis highlights the extensibility of the GeoTox framework to characterize risk at a precise biological level such as the molecular initiating events or key events quantified by curated in-vitro assays.

Second, we demonstrate a new advancement in the GeoTox framework that allows the incorporation of multiple assay end-points into the geospatial risk mapping workflow. Utilizing the **GeoTox** package, we characterize the county-level, annual air toxic carcinogenic risk based on 200+ assays from the key characteristics of carcinogens (KCC) [33]. The assays included are based on the following KCC modes of action: KCC2, Genotoxic Effects; KCC4, Epigenetic Alterations; KCC5, Oxidative Stress; KCC6, Chronic Inflammation; KCC8, Receptor Mediated Effects; and KCC10, Cell Proliferation/Death/Energetics. **GeoTox** incorporates another level of summarization at the assay level such that multi-assay risk is quantified as the "p total quantile of the q assay-level quantiles".

Lastly, we demonstrate the extensibility of **GeoTox** to analyze and visualize individual-level risk that may vary across populations. Population variability in **GeoTox** is based upon the inter-individual variability introduced in the **httk** package to accomodate individual varibility in toxicokinetic processing via Monte Carlo simulations[23]. Here, we visualize the individual variability of multiple-assay based GeoTox risk. We discuss how human experimental or epidemiological data can be used to characterise risk based on the GeoTox framework.

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

3 Results

3.1 Case Study 1: State-wide results from a single in vitro assay

Here, we present county-level, single-assay risk mapping results based on the H2AX histone modification genotoxicity assay (assay id: TOX21 H2AX HTRF CHO Agonist ratio), a potential in-vitro replacement for cancer relative potency factors. All three of the case study results include the intersection of chemicals available in the geospatial exposures, assay information, and valid toxicokinetic parameters. The dose-response results in figure 4A show 7 chemicals with valid hit-calls (i.e. statistically significant departure from null response) to the H2AX histone modification assay and valid toxicokinetic parameter information, which is easily produced with a package call to **plot(geoTox, type = "hill")**. Figure 4B shows the GCA median response mapped to the county level, which is produced with the default call to **plot(geoTox)**. Median GCA mixture responses show risk to these 7 chemicals and this potential cancer target is low when compared to the range of the dose-response functions in figure 4A. The highest risk areas are generally rural counties along the Interstate-95 corridor and near the Virginia border, which indicates a rural confluence of volatile, gas-phase air pollutant exposure and population toxicokinetic characteristics. It is important to note that the chemical space available in this analysis is a small fraction of the overall exposomic chemical space. Lastly, Figure 4C highlights the simplicity of analyzing sensitivity analysis results with a call to **plot(geoTox, type = "sensitivity")**. This analysis highlights the ease of reproducing the analysis from Eccles et al. [12] and producing plots for evaluating chemical response, geospatially mapped risk, and parameter sensitivity.

3.2 Case Study 2: State-wide results from multiple assays

New advancements in the GeoTox framework and **GeoTox** package allow the incorporation of multiple assay end-points into the geospatial risk mapping work-flow, which facilitate a more complete mechanistic risk mapping by incorporating multiple biological end-points or modes of action. Figure 5 shows the summarized risk, quantified by the GCA-based hazard quotient (GCA.HQ.10), for all key characteristics of carcinogens (KCC). The rows show the assay-level summarization by the 10, 50, and 90 quartiles (A-10, A-Med, A-90), respectively. Assay-level calculations are summaries over the individual-level metrics (e.g. hazard quotient or assay response) simulated via Monte Carlo sampling from the **httk** package. The columns show the total summarization, which is the p-th quantile of the assay-level quantiles. This represents a summarization of multiple end-points or modes of action. Here, we chose fifth and tenth quantiles of the hazard quotient as they are conservative risk estimates that avoid the lowest point of departures which are likely capturing cytotoxicity and not the true mode of action [34]. Multi-assay analysis also allows investigation of risk by mode of action or adverse outcome pathways. Figures S1 and S2 in the supporting information show the muti-assay summaries by the KCC for genotoxicity and oxidative stress, respectively. Figure S3 shows the kernel densities of hazard

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

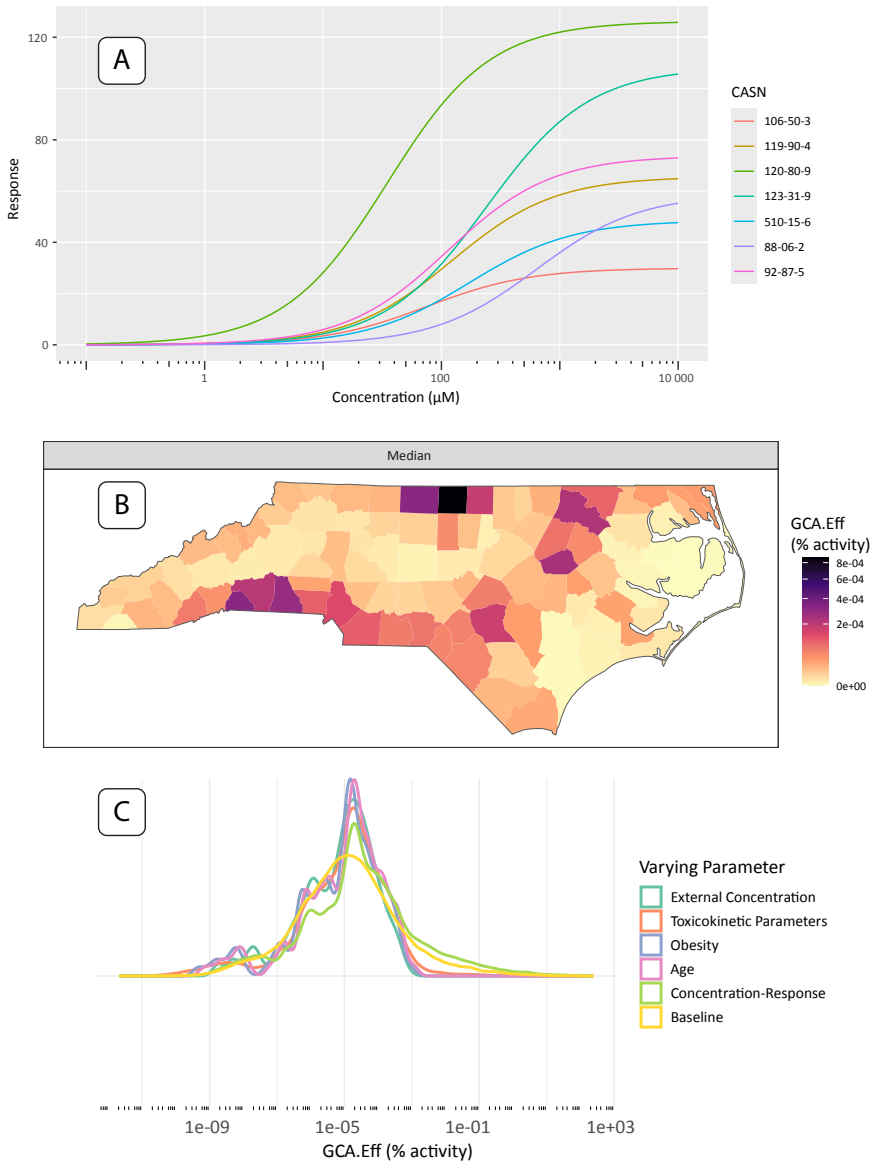

Fig. 4: Single assay results and plots produced by the **GeoTox** package. (A) Fitted dose-response plots for the 2-parameter hill model of the 7 chemicals with hit-calls on the H2AX histone modification genotoxicity assay. (B) Median generalized concentration addition (GCA) response or effect (GCA-Eff, units = percent activity) mapped to the county level. (C) Sensitivity analysis results shown as smoothed kernel densities by analysis parameter.

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

quotients for each assay where each assay is grouped by the KCC mode of action. This analysis highlights the extensibility of the GeoTox framework and the new development of the **GeoTox** package by demonstrating mechanistically-informed risk mapping from multiple modes of action. For example, further investigations into mechanistically-informed risk can be easily conducted via the **GeoTox** package by visualizing combinations of the assay-level and total summarization by adjusting the **assay_quantiles** and **summary_quantiles** parameters, respectively.

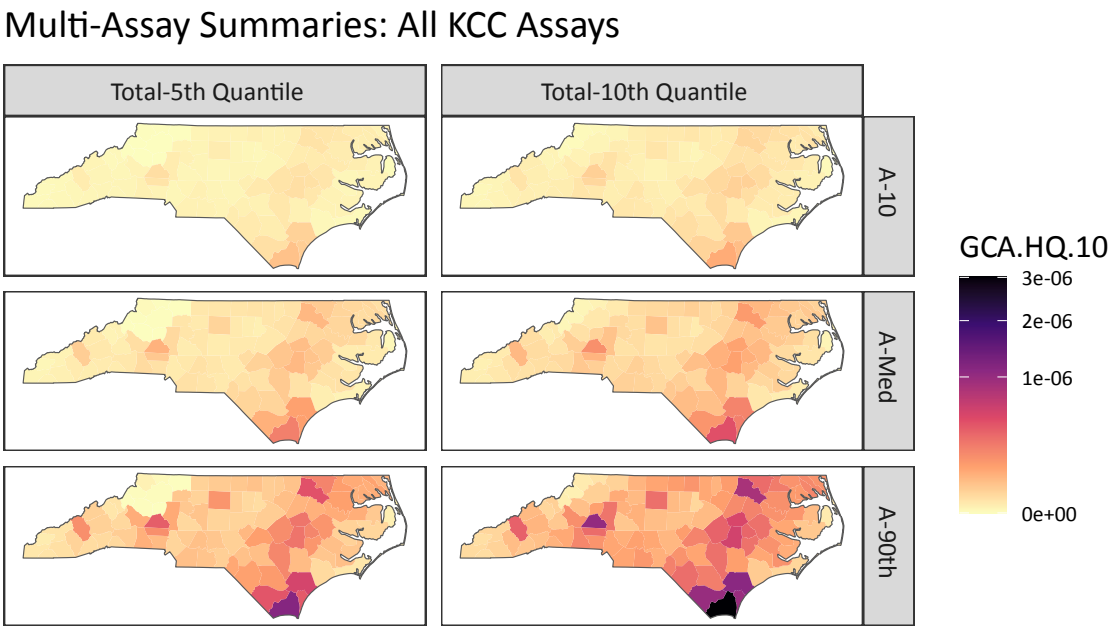

Fig. 5: Multi-assay risk summaries mapped across North Carolina counties using the GCA based hazard quotient. Rows are the individual assay level summarization of 10th, 50th, and 90th quantiles, respectively. The columns are the total summarization as the 5th and 10th quantile, respectively.

3.3 Case Study 3: Characterization of population variability in risk estimated from diverse individuals

The **GeoTox** package is extensibile for estimating and geospatially mapping individual-level risk from chemical mixtures that manifests through individual toxicokinetic variability. This case study highlights how individual toxicokinetic characteristics give rise to the estimated population risks. Figure 6 highlights three North Carolina counties from figure 5 that nominally represent high (Brunswick County, FIPS = 31019), medium (Wake County, FIPS = 37183), and low (Ashe County, FIPS = 37009) risk from the multi-assay risk summarization based on the median or

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

90th quantile assay and 10th quantile total summarization. We show the individual level metrics as jittered points inside kernel density estimates of the hazard quotient, grouped by weight status (i.e. normal vs obese), county, and by the H2AX histone modification genotoxicity assay and multi-assay metrics. In the high risk county, we see that normal weight individuals' multi-assay hazard quotient is shifted upwards compared to the medium and low counties. Additionally, the multi-assay hazard quotient for obese individuals in the high risk county has more variability (i.e. a broader, flatter density) than the obese individuals in the medium and low risk counties. In this example, individuals are simulated from underlying populations, but real-world epidemiological cohort data can used as long as geospatial location information and toxicokinetic parameters are known for the individuals. If individual-level spatial information is available such as a residential addresses, then that information can be added to the GeoTox object through the **set_boundaries()** function where the **region** parameter is an **sf** point object.

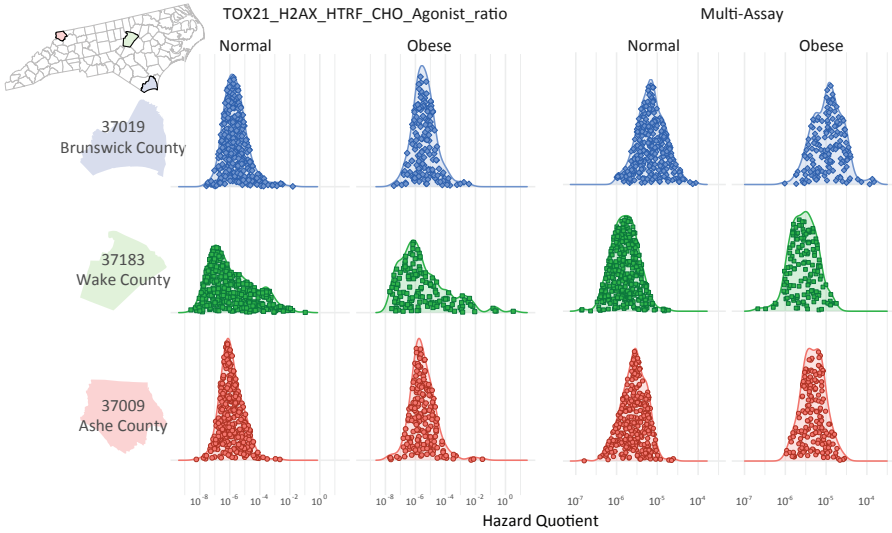

Fig. 6: Kernel densities of the hazard quotient responses grouped by obesity status and assay response. The jittered dots represent the individual responses. The rows are three North Carolina counties nominally representing (descending) high, medium, and low risk, respectively.

4 Discussion

We present the **GeoTox** package for source-to-outcome-continuum modeling that integrates geospatial exposure data, toxicokinetic models, and curated high-throughput screening data in a modular, interconnected, and reproducible manner.

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

The package accommodates individual and population level geospatial and toxicokinetic information. The modularity is intentional, so that users can study factors impacting their step(s) of interest. Further, this provides for integration with key models, such as **httk**, alternative dose-response estimation approaches [27, 35], and geospatial exposure models, as they continue to advance.

The **GeoTox** package implementation was designed to facilitate analysis of population-level variability and susceptibility to adverse effects from combined exposures to chemical mixtures. The population-level parameters impacting risk are bestowed through individual-level simulations from the **httk** package. As signals of variability can arise from several factors, **GeoTox** includes explicit parameters for factors including genetics (*simulated css*), life stage (*age*), physiology (*inhalation rate*), concurrent stressors (*exposure*), and co-morbidities (*obesity*). For complete analysis, provisions for sensitivity analysis of all parameters are also built-in. This allows comparison of simulated scenarios with parameters based on real (i.e. epidemiologically studied) populations.

New research will naturally expand and improve upon the parameters that impact the accuracy and uncertainty in **GeoTox**. For instance, improvements in chemical specific toxicokinetics will lead to comparable improvements in the geospatially mapped risk estimates. Additionally, expanding the chemical space for geospatial exposure maps will complete the exposomic landscape that the framework and package can reasonably capture. The current parameters impacting individual variablity in toxicokinetic processing are age, sex, and obesity. However, it was known that individual genetics such as single-nucleotide polymorphisms (SNPs) can impact toxicokinetic processing. For example, aldehyde dehydrogenase 2 (ALDH2) and aldehyde dehydrogenase 1B (ADH1B) are known variants that greatly impact alcohol metabolism and occurs often in East Asian ancestry [36]. Ginsberg et al. [37] describe six polymorphic enzymes that impact xenobiotic metabolism, which could impart individual-level variability in dose-response to environmental pollutants. Ford et al. [38] have proposed in-vitro NAM for quantifying individual-level variability to chemical responses and highlight the potential for single variants genome wide to impact the individual-level toxicity. If it is known how these variants impact toxicokinetic parameters such as hepatic clearance, then it will endow the framework with the ability to capture individual SNP related differences in risk.

With the introduction of multi-assay or multiple mechanistic endpoint analysis, **GeoTox** will support next-generation risk assessment (NGRA) approaches that utilize NAMs to reduce our reliance on in-vivo animal models, but still incorporate multiple-levels of biological activity and function. Our case study on mapping the total risk quantified by the KCC modes of action provides a straightforward approach to evaluate risk using NAMs that includes multiple routes or modes of toxicity as recommended by Schmeisser et al. [34]. Moreover, our analysis by each KCC category shows that population-level risk can vary based on the mechanistic endpoint. The simplicity of the analysis and plotting functions allow further exploration of causes and policies around mode-of-action variability in population-level chemical responses. Future NGRA developments that can be integrated into **GeoTox** include risk metrics that utilize a more complete or sophisticated mechanistic pathway information.

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Arguably the most important aspect of **GeoTox** is our adherence to FAIR practices in software development that will enable an extensible, reproducible, documented, and maintained resource for the geospatial risk assessment community. **GeoTox** was developed and is maintained according to test-driven development principles, with a series of unit and integration tests ensuring that each function operates correctly on its own and within the cumulative workflow. With adherence to FAIR data principles **GeoTox** is a tested, reliable, and accessible tool which aims to reduce the barriers associated with integrating geospatial information into NGRA or providing geospatial analysts with news tools to integrate health analysis into their previously pure geospatial analyses. As with any sound science or software, minor fixes and methodological improvements will be incorporated through the continuous integration and development process. We provide the user community the forum to report problems and request new features through the code repository (https://github.com/NIEHS/GeoTox/issues).

5 Conclusion

We present **GeoTox**, an open-source, tested, reproducible, and extensible software designed for source-to-outcome-continuum modeling. The software provides researchers and practitioners with a tool for integrating geospatial information into NGRA. Our case studies demonstrate its functionality for geospatially mapping the combined chemical mixture risk quantified by (1) a single assay, AOP key event, or mechanistic endpoint, (2) multiple assay response or disease outcome mode of actions, and (3) individual-level assay-based dose-response. The software was designed with modularity and extensibility such that users can utilize any aspect of the workflow or extend its capabilities for addressing novel environmental, spatial, and human health research questions. We believe there is vast potential to utilize **GeoTox** to elucidate novel genomic and environmental risk questions at both the population and individual level. And most critically, the software is developed with computational best-practices such that it will me maintained and evolve with the needs of the community in addressing complex human health problems.

6 List of Abbreviations

- **AEP** Aggregate Exposure Pathway
- **AOP** Adverse Outcome Pathway
- **cHTS** curated high-throughput screening
- **GCA** Generalized Concentration Addition
- **httk** high-throughput toxicokinetics
- **KCC** Key Characteristics of Carcinogens
- **NGRA** Next Generation Risk Assessment
- **SNP** Single Nucleotide Polymorphism

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Acknowledgements

This work is supported by the National Institute of Environmental Health Sciences, Division of Translational Toxicology, Division of Intramural Research, and the Spatiotemporal Exposures and Toxicology group under project number ZIA ES103368-02.

References

[1] Environmental Health Sciences, N.I.: 2018 - 2023 Strategic Plan: Advancing Environmental Health Sciences, Improving Health. Technical report, US Department of Health and Human Services (2018)

[2] Vermeulen, R., Schymanski, E.L., Barabási, A.-L., Miller, G.W.: The exposome and health: Where chemistry meets biology. Science **367**(6476), 392–396 (2020)

[3] Wild, C.P.: The exposome: from concept to utility. International journal of epidemiology **41**(1), 24–32 (2012)

[4] Wild, C.P., Scalbert, A., Herceg, Z.: Measuring the exposome: a powerful basis for evaluating environmental exposures and cancer risk. Environmental and molecular mutagenesis **54**(7), 480–499 (2013)

[5] Walker, D.I., Valvi, D., Rothman, N., Lan, Q., Miller, G.W., Jones, D.P.: The metabolome: A key measure for exposome research in epidemiology. Current Epidemiology Reports **6**, 93–103 (2019)

[6] Miller, G.W.: Integrating the exposome into a multi-omic research framework. Exposome **1**(1), 002 (2021)

[7] Rappaport, S.M., Smith, M.T.: Environment and disease risks. science **330**(6003), 460–461 (2010)

[8] Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., *et al.*: Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environmental Toxicology and Chemistry: An International Journal **29**(3), 730–741 (2010)

[9] Teeguarden, J.G., Tan, Y.M., Edwards, S.W., Leonard, J.A., Anderson, K.A., Corley, R.A., Kile, M.L., Simonich, S.M., Stone, D., Tanguay, R.L., Waters, K.M., Harper, S.L., Williams, D.E.: Completing the Link between Exposure Science and Toxicology for Improved Environmental Health Decision Making: The Aggregate Exposure Pathway Framework. Environmental Science and Technology **50**(9), 4579–4586 (2016) https://doi.org/10.1021/acs.est.5b05311 PMCID: PMC4854780. ISBN: 0013-936X _eprint: 15334406

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

[10] Hines, D.E., Conolly, R.B., Jarabek, A.M.: A quantitative source-to-outcome case study to demonstrate the integration of human health and ecological end points using the aggregate exposure pathway and adverse outcome pathway frameworks. Environmental Science & Technology **53**(18), 11002–11012 (2019)

[11] Price, P.S., Jarabek, A.M., Burgoon, L.D.: Organizing mechanism-related information on chemical interactions using a framework based on the aggregate exposure and adverse outcome pathways. Environment international **138**, 105673 (2020)

[12] Eccles, K.M., Karmaus, A.L., Kleinstreuer, N.C., Parham, F., Rider, C.V., Wambaugh, J.F., Messier, K.P.: A geospatial modeling approach to quantifying the risk of exposure to environmental chemical mixtures via a common molecular target. Science of The Total Environment **855**, 158905 (2023)

[13] USEPA: Exposure factors handbook 2011 edition (final). Washington, DC (2011)

[14] Pearce, R.G., Setzer, R.W., Strope, C.L., Sipes, N.S., Wambaugh, J.F.: httk: R package for high-throughput toxicokinetics. Journal of Statistical Software **79**(4), 1–25 (2017) https://doi.org/10.18637/jss.v079.i04 PMC6134854

[15] Daniel, A.B., Choksi, N., Abedini, J., Bell, S., Ceger, P., Cook, B., Karmaus, A.L., Rooney, J., To, K.T., Allen, D., *et al.*: Data curation to support toxicity assessments using the integrated chemical environment. Frontiers in Toxicology **4**, 987848 (2022)

[16] Loewe, S.t., Muischnek, H.: Über kombinationswirkungen: Mitteilung: Hilfsmittel der fragestellung. Naunyn-Schmiedebergs Archiv für experimentelle Pathologie und Pharmakologie **114**, 313–326 (1926)

[17] Howard, G.J., Webster, T.F.: Generalized concentration addition: a method for examining mixtures containing partial agonists. Journal of theoretical biology **259**(3), 469–477 (2009)

[18] Price, P.S.: The hazard index at thirty-seven: New science new insights. Current Opinion in Toxicology **34**, 100388 (2023)

[19] Hester, J., Angly, F., Hyde, R., Chirico, M., Ren, K., Rosenstock, A., Patil, I.: Lintr: A 'Linter' for R Code. (2024). R package version 3.1.2, https://lintr.r-lib.org. https://github.com/r-lib/lintr

[20] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H.: Welcome to the tidyverse. Journal of Open Source Software **4**(43), 1686 (2019) https://doi.org/10.21105/joss.01686

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

[21] Wickham, H.: testthat: Get started with testing. The R Journal **3**, 5–10 (2011)

[22] Wickham, H.: Advanced R, 2nd Edition. chapman and hall/CRC, ??? (2019)

[23] Ring, C.L., Pearce, R.G., Setzer, R.W., Wetmore, B.A., Wambaugh, J.F.: Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. Environment International **106**, 105–118 (2017)

[24] Ritz, C., Baty, F., Streibig, J.C., Gerhard, D.: Dose-response analysis using r. PLOS ONE **10**(e0146021) (2015)

[25] Filer, D.L., Kothiya, P., Setzer, R.W., Judson, R.S., Martin, M.T.: tcpl: the toxcast pipeline for high-throughput screening data. Bioinformatics **33**(4), 618–620 (2017)

[26] Wheeler, M.W., Blessinger, T., Shao, K., Allen, B.C., Olszyk, L., Davis, J.A., Gift, J.S.: Quantitative risk assessment: Developing a bayesian approach to dichotomous dose–response uncertainty. Risk Analysis **40**(9), 1706–1722 (2020)

[27] Wheeler, M.W., Lim, S., House, J.S., Shockley, K.R., Bailer, A.J., Fostel, J., Yang, L., Talley, D., Raghuraman, A., Gift, J.S., *et al.*: Toxicr: A computational platform in r for computational toxicology and dose–response analyses. Computational Toxicology **25**, 100259 (2023)

[28] Centers for Disease Control and Prevention: PLACES. Accessed: 2024-05-31 (2024). https://www.cdc.gov/places

[29] Office of Air Quality Planning and Standards: Technical support document: Epa's air toxics screening assessment, 2020 airtoxscreen tsd. Technical report, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711 (February 2024). EPA's Air Toxics Screening Assessment

[30] Williams, A.J., Lambert, J.C., Thayer, K., Dorne, J.-L.C.: Sourcing data on chemical properties and hazard data from the us-epa comptox chemicals dashboard: A practical guide for human risk assessment. Environment international **154**, 106566 (2021)

[31] National Toxicology Program's Integrated Chemical Environment: Data retrieved by https://ice.ntp.niehs.nih.gov/api/v1/search. https://ice.ntp.niehs.nih.gov/. Released v4.0, released March 2023, retrieved on June 1, 2024.

[32] Dreij, K., Mattsson, Å., Jarvis, I.W., Lim, H., Hurkmans, J., Gustafsson, J., Bergvall, C., Westerholm, R., Johansson, C., Stenius, U.: Cancer risk assessment of airborne pahs based on in vitro mixture potency factors. Environmental science & technology **51**(15), 8805–8814 (2017)

[33] Smith, M.T., Guyton, K.Z., Gibbons, C.F., Fritz, J.M., Portier, C.J., Rusyn,

medRxiv preprint doi: https://doi.org/10.1101/2024.09.23.24314096; this version posted November 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

I., DeMarini, D.M., Caldwell, J.C., Kavlock, R.J., Lambert, P.F., *et al.*: Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis. Environmental health perspectives **124**(6), 713–721 (2016)

[34] Schmeisser, S., Miccoli, A., Bergen, M., Berggren, E., Braeuning, A., Busch, W., Desaintes, C., Gourmelon, A., Grafström, R., Harrill, J., *et al.*: New approach methodologies in human regulatory toxicology–not if, but how and when! Environment International **178**, 108082 (2023)

[35] Zilber, D., Messier, K.: Reflected generalized concentration addition and bayesian hierarchical models to improve chemical mixture prediction. Plos one **19**(3), 0298687 (2024)

[36] Cho, Y., Lin, K., Lee, S.-H., Yu, C., Valle, D.S., Avery, D., Lv, J., Jung, K., Li, L., Smith, G.D., *et al.*: Genetic influences on alcohol flushing in east asian populations. BMC genomics **24**(1), 638 (2023)

[37] Ginsberg, G., Smolenski, S., Neafsey, P., Hattis, D., Walker, K., Guyton, K.Z., Johns, D.O., Sonawane, B.: The influence of genetic polymorphisms on population variability in six xenobiotic-metabolizing enzymes. Journal of Toxicology and Environmental Health, Part B **12**(5-6), 307–333 (2009)

[38] Ford, L.C., Jang, S., Chen, Z., Zhou, Y.-H., Gallins, P.J., Wright, F.A., Chiu, W.A., Rusyn, I.: A population-based human in vitro approach to quantify inter-individual variability in responses to chemical mixtures. Toxics **10**(8), 441 (2022)