HHS Public Access

Author manuscript

*Sci Total Environ*. Author manuscript; available in PMC 2024 January 10.

Published in final edited form as:

*Sci Total Environ*. 2023 January 10; 855: 158905. doi:10.1016/j.scitotenv.2022.158905.

# A geospatial modeling approach to quantifying the risk of exposure to environmental chemical mixtures via a common molecular target

**Kristin M. Eccles**[a], **Agnes L. Karmaus**[b], **Nicole C. Kleinstreuer**[a], **Fred Parham**[a], **Cynthia V. Rider**[a], **John F. Wambaugh**[c], **Kyle P. Messier**[a,*]

[a]National Institute of Environmental Health Science, Division of the Translational Toxicology, Durham, USA

[b]Integrated Laboratory Systems, an Inotiv Company, Morrisville, NC, USA

[c]United States Environmental Protection Agency, Center for Computational Toxicology and Exposure, Durham, USA

## Abstract

In the real world, individuals are exposed to chemicals from sources that vary over space and time. However, traditional risk assessments based on in vivo animal studies typically use a chemical-by-chemical approach and apical disease endpoints. New approach methodologies (NAMs) in toxicology, such as in vitro high-throughput (HTS) assays generated in Tox21 and ToxCast, can more readily provide mechanistic chemical hazard information for chemicals with no existing data than in vivo methods. In this paper, we establish a workflow to assess the joint action of 41 modeled ambient chemical exposures in the air from the USA-wide National Air Toxics Assessment by integrating human exposures with hazard data from curated HTS (cHTS) assays to identify counties where exposure to the local chemical mixture may perturb a common biological target. We exemplify this proof-of-concept using CYP1A1 mRNA up-regulation. We first estimate internal exposure and then convert the inhaled concentration to a steady state plasma concentration using physiologically based toxicokinetic modeling parameterized with county-specific information on ages and body weights. We then use the estimated blood plasma concentration and the concentration-response curve from the in vitro cHTS assay to determine the



*Corresponding author at: National Institute of Environmental Health Sciences, 530 Davis Dr., Keystone Building, Durham, NC 27713, USA. kyle.messier@nih.gov (K.P. Messier).

Disclaimer

The United States Environmental Protection Agency (EPA), through its Office of Research and Development (ORD), funded the research described here. The views expressed in this publication are those of the authors and do not necessarily represent the views or policies of the U.S. EPA. Reference to commercial products or services does not constitute an endorsement.

CRediT authorship contribution statement

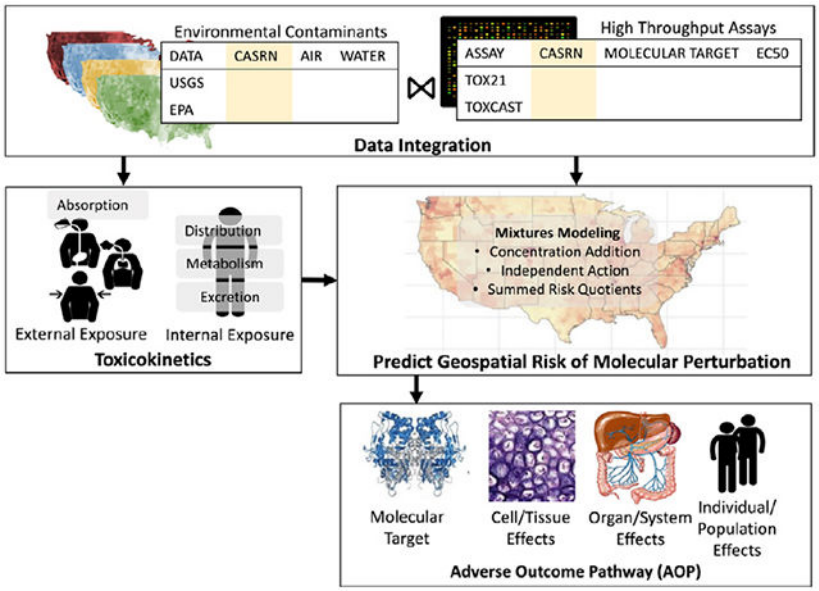
**Kristin M. Eccles:** Conceptualization, Methodology, Formal analysis, Writing – original draft. **Agnes L. Karmaus:** Methodology, Writing – review & editing. **Nicole C. Kleinstreuer:** Methodology, Writing – review & editing. **Fred Parham:** Methodology, Writing – review & editing. **Cynthia V. Rider:** Methodology, Writing – review & editing. **John F. Wambaugh:** Methodology, Writing – review & editing. **Kyle P. Messier:** Conceptualization, Methodology, Formal analysis, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

chemical-specific effects of the mixture components. Three mixture modeling methods were used to estimate the joint effect from exposure to the chemical mixture on the activity levels, which were geospatially mapped. Finally, a Monte Carlo uncertainty analysis was performed to quantify the influence of each parameter on the combined effects. This workflow demonstrates how NAMs can be used to predict early-stage biological perturbations that can lead to adverse health outcomes that result from exposure to chemical mixtures. As a result, this work will advance mixture risk assessment and other early events in the effects of chemicals.

## Graphical Abstract




### Keywords

Mixtures; High-throughput assay; New approaches methodologies; Exposome; Spatial risk assessment; Dose addition

## 1. Introduction

Modern environments contain diverse and heterogeneously distributed toxicants resulting in localized and unique exposures to complex mixtures for individuals. To characterize the human health risk from such chemical exposures, an assessment requires information on the potential for the chemical to cause an adverse effect (i.e., hazard), the degree of response at differing exposure levels (i.e., concentration-response), and the concentration of a chemical that an individual experiences (i.e., exposure) (National Research Council, 1983). Characterizing risk from environmental mixtures of chemicals is particularly challenging because hazard, concentration-response, and exposure data are required for all relevant chemicals, and challenges integrating data across chemicals limit their use in human health risk assessment.

Many environmental chemicals lack traditional chemical hazard information due to the time-consuming nature of in vivo animal studies (Judson et al., 2009). The introduction and use of new approach methodologies (NAMs) for toxicology fill this data gap in hazard assessment using in vitro, in silico, and chemoinformatic methods to rapidly inform chemical hazard characterization (Kavlock et al., 2018; Zavala et al., 2020). Specifically, in vitro high-throughput screening (HTS) assays have been used to generate concentration-response information for >10 K chemicals and ~ 3000 different assay endpoints that inform on a variety of biochemical and pathway targets, including nuclear receptors, cytotoxicity, cell stress, etc. (Krewski et al., 2009; Richard et al., 2016; Sipes et al., 2017). This information has been used to prioritize chemicals for risk assessment and link chemical exposures to perturbations in biological pathways (e.g., Auerbach et al., 2016; Judson et al., 2015; Judson et al., 2011; Kleinstreuer et al., 2017). More recently, NAMs have also been used to estimate the risk of environmental exposures, such as environmental water samples (Escher et al., 2018; Neale et al., 2020; Wambaugh et al., 2019). However, contextualizing NAM-based chemical hazard characterization with environmentally relevant concentrations and placing that information into a human-relevant biological context remains a challenge.

Another data gap is the accurate determination of human exposure to multiple chemicals from environmental sources such as ambient air. Advancements in methods used to determine human exposure to environmental chemicals include biomonitoring and environmental modeling techniques such as land-use regression and geostatistical models (Dennis et al., 2017; Katzfuss, 2017; Katzfuss et al., 2020; Messier et al., 2014), mechanistic chemical transport and dispersion models (Appel et al., 2021), and hybrid mechanistic and geostatistical models (Cleland et al., 2020; Messier and Katzfuss, 2021; Wang et al., 2016). Additionally, exposure models have been used to estimate simple exposure heuristics of near-field (i.e., personal and consumable product) and far-field (i.e., ambient environment) chemical sources (Wambaugh et al., 2014) and can estimate exposure and internal concentration by chemical class and real-world demographic characteristics (e.g., age, sex). While NAMs and simulation-based fate and transport models have been useful for prioritizing chemicals with high exposure and high potency, they have not yet considered the geospatial and temporal distributions of chemicals that contribute to the population variability among human exposures. For example, work by Ring et al. (2017) and Ring et al. (2019) uses HTS hazards and estimated exposure to prioritize the risk of environmental chemicals, however, this has only been completed at the USA level.

Traditional risk assessments are typically completed on a chemical-by-chemical basis (Fay et al., 2018; Hsieh et al., 2021). Therefore, the true impact of chemicals may be greater than predicted by even the most conservative of individual chemical risk assessments due to the potential for additive or synergistic effects (Kortenkamp and Faust, 2018). Chemicals that are individually below toxicity thresholds, such as the No-Observed-Adverse-Effect-Level (NOAEL), and exert a common effect, have been shown to additively combine to harmful levels (Kortenkamp et al., 2007; Kortenkamp and Faust, 2018).

Several frameworks have been proposed to address these data and knowledge gaps in environmental health research, including the exposome and eco-exposome. The exposomics field broadly seeks to estimate the totality of chemical and non-chemical exposures

over a lifetime (Miller and Jones, 2014; Vermeulen et al., 2020; Wild, 2005). However, individual variability in chemical exposure and responses due to chemical toxicokinetics, analytical methods for detecting complex mixtures of exogenous chemicals and endogenous metabolites, and attributing complex spatial and temporal distributions of chemicals to individual biological measurements still limit the field of exposomics (Escher et al., 2020c; Scholz et al., 2022; Zhu et al., 2021). The eco-exposome framework uses a holistic approach to link environmental chemical sources to an adverse outcome utilizing the existing Aggregate Exposure Pathway (AEP; Tan et al., 2018; Teeguarden et al., 2016) and Adverse Outcome Pathway (AOP; Ankley et al., 2010) frameworks to provide a modular pathway-based method for tracking chemicals from source through exposure through a biological mechanism to an adverse outcome (Hines et al., 2018; Scholz et al., 2022; Tan et al., 2018). The AEP framework quantifies the events that make up the fate and transport process of environmental chemicals through different media and, when appropriate, chemical transformations that occur in the environment. The framework also addresses the events in the exposure of the individual and the processes of adsorption, distribution, metabolism, and excretion that determine the relationship between the external dose and internal concentration of a chemical (or its active metabolites). These internal concentrations are referred to as the Target Site Exposures (TSEs) and are defined as the relevant exposures for a specific Molecular Initiating Event (MIE) (Teeguarden et al., 2016). The AOP framework defines an AOP as the set of events that relate TSEs to MIEs and ultimately to an adverse outcome through a cascade of pathway-specific biological key events that occur at multiple levels of biological organization (Ankley et al., 2010). Multiple AOPs with common events, MIEs, or adverse outcomes can be combined to form AOP networks (Villeneuve et al., 2018). The combination of an AOP network linking one or more MIEs to an adverse outcome with the AEP generated TSEs provide a tractable approach to comprehensively track and quantify the combined effects of multiple chemicals over the source-exposure-outcome continuum (Hines et al., 2019; Price et al., 2020; Scholz et al., 2022).

The objective of this research is to demonstrate a workflow for characterizing the geospatial risk of perturbing molecular targets that are implicated in adverse human health outcomes based on exposure to spatially explicit chemical mixtures. The combined AEP-AOP is linked to geospatially referenced estimates of exposures to multiple chemicals from environmental sources with chemical-specific hazard information, e.g., from HTS assays, providing a mechanism-based approach for assessing geospatial variation in risk from real-world exposures to chemical mixtures. Using a geospatially resolved AEP-AOP framework can further enrich what is known about localized chemical exposure, support the development of AEPs, and provide a method for including environmentally relevant chemical exposures that can be connected to AOPs. Further, as HTS assays can inform on MIE and other key biological events within an AOP, a bottom-up approach using exposure concentrations of environmentally relevant chemical mixtures known to act on shared molecular targets can provide the opportunity to better understand mechanisms of disease progression (Fay et al., 2018; Kleinstreuer et al., 2016). By focusing on the combined chemical exposures' effects on early events in AEP-AOP networks (as measured by HTS assays), the approach used in this paper will provide information on the mechanisms of

chemical interactions that can inform on combined risk. The toxicological response assessed is a perturbation of a molecular target that is an early event in AEP-AOP networks.

## 2. Methods

### 2.1. Overview of workflow

Here we provide a general overview of the workflow that integrates a set of modeled environmental chemical exposure data with cHTS assay data to identify chemicals that co-occur in the environment and perturb the same molecular target. Fig. 1 supplements this general overview by linking equations and specific details of the proof-of-concept example found in the methods. First, we selected the cHTS assay that informs on the molecular target of interest and determined the overlap between the chemicals that activate the assay and the external concentration data (e.g., geospatial models and personal monitoring data). We then estimate the internal concentration for each of the identified chemicals based on the external concentration using the corresponding equation appropriate for the route of exposure from the EPA Exposure Factors Handbook (U.S. EPA, 2011). Then, we convert these internal concentrations to an equivalent steady-state plasma concentrations in adults using the physiologically based toxicokinetic (PBTK) modeling in *httk* (Breen et al., 2021; Pearce et al., 2017a). This modeling u ses county-specific demographic information (i.e., age and obesity status) to estimate the variation in physiological characteristics that influence the toxicokinetic model parameters to generate a scaling factor for each chemical, which is applied to the internal concentration. These equivalent steady-state plasma concentrations for each chemical are then combined into one predicted exposure-response for the mixture using mixtures modeling methods (i.e., concentration addition, response addition, and a summed risk quotient). These methods utilize parameters from the concentration-response curves in the cHTS data. The result is a set of estimates of the risk of the combined effect of the chemicals on the HTS endpoint for the average individual in each US county, which is then mapped. Finally, we perform a Monte Carlo uncertainty analysis to quantify the variability in the input parameters into the workflow, including external chemical concentration, the population parameters of age and obesity status used in the PBTK model, the concentration-response parameter using an $EC_{50}/EC_{10}$, and top of the curve.

### 2.2. Data

**2.2.1. Population demographics—**The spatial county-level, ($\boldsymbol{s}$), population demographic distributions, $P(\boldsymbol{s})$, of age were obtained from the US Census. These data were aggregated from 18 groups into 9 equal intervals between groups ages between 0 and > 85 years old, however, we only use the groups that are ≥ 18 to match the obesity data. The county-level percentages and standard deviations of obesity were obtained from *the Centers for Disease Control PLACES data* (Centers for Disease Control and Prevention, 2021). *Obesity percentage is* calculated as the percentage of people ≥18 years who have a body mass index (BMI) ≥ 30.0 $kg/m^2$ calculated from self-reported weight and height divided by the total number of respondents. This percentage excludes respondents measuring <3 f. or ≥ 8 ft, weighing <50 lbs. or ≥ 650 lbs., BMI of <12 $kg/m^2$ or ≥100 $kg/m^2$, and pregnant individuals. Both age and obesity status are important factors in determining the inhalation rate, which affects the inhaled chemical dose and governs the chemical's metabolic clearance

and the chemical's resulting internal dose (Fig. 1). Rates of inhalation are higher in younger individuals than older individuals, and obese individuals have a higher inhalation rate than normal-weight individuals (U.S. EPA, 1992). Higher rates of inhalation results in a higher inhaled chemical dose. Toxicokinetic parameters such as metabolic clearance are higher in younger individuals than in older individuals, and individuals with normal weight have higher metabolic clearance of chemicals than obese individuals (Ring et al., 2017). High metabolic clearance results in lower internal chemical concentrations.

**2.2.2. Ambient chemical concentrations—**The Environmental Protection Agency's 2014 National Air Toxics Assessment (NATA) modeled environmental concentrations for 177 individual chemicals and 11 chemical groups (e.g., polycyclic aromatic hydrocarbons) using the combined American Meteorological Society/Environmental Protection Agency Regulatory Model (AERMOD) and EPA's Community Multiscale Air Quality (CMAQ) Modeling System version 5.2, which use atmospheric dispersion and chemical transport models, respectively. Inputs into these models include information from the National Emissions Inventory (NEI), including point source data (e.g., industrial emissions, gas stations, agricultural and livestock emissions), nonpoint source emissions (e.g., residential heating, fugitive dust), and vehicle emission (U.S. EPA, 2018). The NATA dataset estimates ambient air concentrations of chemicals at a census block level. Achieving this level of resolution in contaminant estimates has been a limitation of platforms such as USEtox, which estimates exposures at the continental and global scales (Huang et al., 2021). We aggregated the concentration to a county-level to match the resolution of the population demographic data (e.g., obesity data is only available at the census tract level). To do this, we calculate the mean of the census blocks within the county. We obtain the county distributions for the external air concentration, $C_{ext}(\boldsymbol{s})$ using the standard deviation of the mean.

**2.2.3. Curated concentration-response curves from high-throughput screening assays—**Curated concentration-response curves from high-throughput screening (cHTS) assays originating from the Tox21 and ToxCast HTS programs were obtained from the National Toxicology Program's Integrated Chemical Environment (Abedini et al., 2021). In brief, the assay data are first modeled as concentration-response curves using the tcpl package in R by the US EPA (Filer et al., 2017) and then are subject to quality assurance/quality control curation (Abedini et al., 2021). We use the curated hit calls (i.e., indicating biochemical activity in an assay) and raw concentration-response data. The raw concentration-response data are refit to a two-parameter hill model (Eq. (1)), which assumes a slope of 1 to be more amenable to mixtures modeling (see below):

$$u_i = \frac{tp}{1 + 10^{(ga - x_i)}} \tag{1}$$

where, $u_i$ is the response, $tp$ is the maximal response at the top of the curve, $ga$ is the curve $EC_{50}$, and $x_i$ is the chemical concentration. The two-parameter model was fit using the same constraints on the model parameters and algorithm as applied by tcpl (Filer et al., 2017), including defining the log-likelihood as a Student's $t$-distribution with four degrees of freedom (Table S1). The cHTS data for each chemical was linked to the corresponding

ambient chemical concentrations from NATA data using unique CAS Registry Numbers (CASRNs).

**2.2.4. CYP1A1 up-regulation assay—**To demonstrate our approach, we use an assay measuring CYP1A1 mRNA transcription up-regulation (LTEA_HepaRG_CYP1A1_up) (Franzosa et al., 2021). CYP1A1 is one enzyme in the cytochrome P450 family that has many functions, including the metabolism of carcinogens and hormones (Mescher and Haarmann-Stemmann, 2018; Nebert and Dalton, 2006). The induction of CYP1A1 expression can be a biomarker of aryl hydrocarbon receptor (AHR) activation and environmental exposure to polycyclic aromatic compounds (PAHs) and organochlorines (e.g., dioxins) (Mescher and Haarmann-Stemmann, 2018). In humans, the CYP1A1 enzyme is among the most active cytochrome P450 enzymes responsible for metabolizing procarcinogens (e.g., PAHs) (Androutsopoulos et al., 2009). Other chemical exposures metabolized by the CYP1A1 enzyme could further upregulate the mRNA expression (Nebert and Dalton, 2006). This assay informs on the change in the number of CYP1A1 transcripts in HepaRG immortalized human liver cell cultures. The “up” direction indicates increased expression (induction) relative to the control. We chose this assay due to the biological relevance to humans and the high number of active hit calls out of the chemicals tested that were also quantified in NATA ($n$ = 42/177); however, toxicokinetic information was unavailable for one chemical thus the final $n$ = 41. The high number of environmentally relevant chemicals in the mixture highlights the capabilities of our proposed method in this case study, however, the approach is amenable to a wide variety of assays.

## 2.3. Human internal concentration calculation

The mass of ambient chemical in air, $C_{ext}(\boldsymbol{s})(\frac{mg^{air}}{m^3})$, is converted to a mass dose in a human $(\frac{mg}{kg})$ to obtain an internal chemical dose, $D_{int}$, via an inhalation route (U.S. EPA, 1992):

$$D_{int}(\boldsymbol{s}) = \frac{C_{ext}(\boldsymbol{s}) \times IR(\boldsymbol{s}) \times t}{\mathrm{BW}(\mathbf{s})} \quad (2)$$

where, the external air concentration, $C_{ext}(\boldsymbol{s})$ is multiplied by the spatially dependent county inhalation rate, $IR(\boldsymbol{s})$, $\left(\frac{m^3}{day}\right)$, and $t$ is the total time, in our example, of 365 days; and $BW(\boldsymbol{s})$ is the county-level human body weight (in $kg$). We conservatively assume that 100 % of the chemical is absorbed, and therefore the attributable fraction has a value of 1 and is not included in Eq. (2). We also assume that the steady-state concentration of each chemical is achieved in 1 day, and the repeated daily exposure to these chemicals accumulates additively to produce a summed internal concentration (Pearce et al., 2017b). We use this cumulative exposure as the starting point for determining the effect.

The $IR$ and $BW$ distributions are based on population parameters from the PLACES dataset and vary by county. The $IR$ was obtained from the exposure factors handbook (U.S. EPA, 2011) and depended on age and sex, among other intrinsic factors. The inhalation rates for men and women of obese and normal weights were similar, so they were averaged to a

marginal $IR$ per age group. We also match these exposure factors with the country-specific distributions of age and obesity rates. Based on the county, the population demographics (e.g., age and obesity status) and external chemical exposures ($C_{ext}$), $D_{int}$ varies spatially by county. Fig. 1 shows how the population demographics data, $P(\boldsymbol{s})$, inform the county-level dependence of the human internal concentration. Since this example uses annual averages for ambient chemical concentrations, the inhalation dose is assumed to be constant over time. This equation could be replaced with other equations to estimate internal concentration from the EPA Exposures Factors Handbook (U.S. EPA, 2011), thus making the method generalizable for other routes of exposure (i.e., ingestion) if reasonable estimates are available in the spatial domain.

## 2.4. Internal dosimetry

We use county-specific distributions of age and obesity to develop $C_{ss}(\boldsymbol{s})$ scaling factors that are unique for each county and chemical, which are used to convert the inhaled dose to the spatially explicit internal dose $D_{int}(\boldsymbol{s})$ then to an in vitro equivalent blood plasma concentration $C_{plasma}(\boldsymbol{s})$ (Fig. 1). To do this, we used physiologically based toxicokinetic (PBTK) modeling to estimate a steady-state blood plasma concentration based on an inhaled dose of 1 mg/kg/day for each chemical exposure (Sipes et al., 2017). We then use this as a linear scaling factor to convert our inhaled mass concentration to an in vitro equivalent steady-state blood plasma concentration, which allows us to use the bioactivity concentration-response data from the high throughput screening assays, given using Eq. (3) (Breen et al., 2021):

$$C_{plasma}(\boldsymbol{s}) = C_{ss}(\boldsymbol{s}) \times D_{int}(\boldsymbol{s}) \tag{3}$$

where, $C_{plasma}(\boldsymbol{s})$ is the county-specific in vitro equivalent plasma concentration in $\mu$M and $C_{ss}(\boldsymbol{s})$ is the steady-state plasma concentration in $\frac{\mu M}{mg\ /\ kg}$ estimated from *httk,* an R package for high-throughput toxicokinetic modeling, developed using a three-compartment model (Pearce et al., 2017a) and simulating a dose of 1 mg/kg. Population parameters such as obesity status and age are used to parameterize the $C_{SS}$ scaling factor as these parameters affect important toxicokinetics, including the hepatic clearance rate and the fraction of the chemical unbound to plasma protein. Interindividual variability of these factors are considered in *httk* using a Monte Carlo simulation where one average $C_{SS}$ scaling factor is created per chemical using the population parameters from the National Health and Nutrition Examination Survey (NHANES) (Ring et al., 2017).

## 2.5. Mixture modeling and geospatial risk assessment

In the geospatial risk assessment, we present three different but commonly used methods for calculating the combined effects of exposure to chemical mixtures. Two of these methods, concentration addition (CA) and risk quotient (RQ), fall under the concentration addition paradigm, where exposure concentrations are normalized using a reference concentration to predict a combined response or risk. The third mixture method used in this paper is independent action (IA). Each of these methods is described in detail below. For simplicity, we will refer to these methods as CA, IA, and the sum of RQs, in this order, throughout

the paper since the units of CA and IA are both predicted responses (Log2 fold change of CYP1A1 mRNA expression), and the sum of RQs is a unitless measure.

**2.5.1. Concentration addition—**CA is typically used for chemicals that share a common mode of action or molecular target. In this proof-of-concept, all chemicals act on the same molecular target, the upregulation of CYP1A1 mRNA expression, making it an appropriate choice. CA is given by Eq. (4) (Altenburger et al., 2000; Berenbaum, 1985):

$$ECx_{mix}(\boldsymbol{s}) = \left( \sum_{i=1}^{n} \frac{p_i(\boldsymbol{s})}{ECx_i} \right)^{-1} \tag{4}$$

where, $ECx_{mix}(\boldsymbol{s})$ is the county-level effective concentration of the mixture, $p_i(\boldsymbol{s})$ is the proportion of chemical $i$ present in the in vitro steady-state plasma concentration chemical mixture, and $ECx_i$ is the individual chemical concentration and elicits the same response as the mixture. Here we use an $EC_{50}$, the concentration that produces the half-maximal response. However, CA is limited in its ability to predict the joint effect of chemical mixtures when the chemicals have a different maximum response. When this occurs, Generalized Concentration Addition (GCA) methods can be used to relax the assumption of equal concentration-response maxima (Howard et al., 2010; Watt et al., 2016):

$$1 = \Sigma_{i=1}^{n(s)} \frac{C_i(\boldsymbol{s})}{f_i^{-1}(R(\boldsymbol{s}))} \tag{5}$$

where, $C_i(\boldsymbol{s})$ is the concentration of an individual chemical from Eq. (3), $f_i^{-1}$ is the inverse function of the individual concentration-response curves as given in Eq. (1), and $R(\boldsymbol{s})$ is the predicted county-level response. The equation adds to one for mixtures with additive relationships (opposed to synergistic >1, or antagonistic <1), which we assume in this example. While there is an analytical solution for the 2-parameter hill model, we estimate the mixture response (i.e., effect level), $R(\boldsymbol{s})$ using numerical optimization where the objective function is the squared difference between the left and right sides of Eq. (5). This approach is generalizable to other equations that do not have a neat analytical solution making this estimation approach amenable to other concentration-response models (e.g., Weibull) and their respective inverse functions within the standard assumption of GCA.

**2.5.2. Independent action—**IA is typically used to model chemical mixtures when the chemicals have different mechanisms of action. While the chemicals chosen in this example all act on the same molecular target, there is a lot of variability in the concentration-response curves for the individual chemicals, thus the non-uniform curves may act more like independent action. IA is given by Eq. (6) (Backhaus et al., 2000; Watt et al., 2016):

$$R(\boldsymbol{s}) = \alpha_{max}\left(1 - \prod_{i=1}^{n(\boldsymbol{s})}\left(1 - \frac{f_i[X]_i}{\alpha_{max}}\right)\right) \tag{6}$$

where, $R(\boldsymbol{s})$ is the county-level response, $n(\boldsymbol{s})$ is the number of observed chemicals in a given county, $\alpha_{max}$ scales the overall response to the maximum mixture response, and $f_i$ is the individual concentration-response curve as given in Eq. (1) that produces a response at

concentration [$X$], and the unit is the same as the response unit of the assay, which in this example is Log2 fold change in mRNA expression.

**2.5.3. Risk quotient—**To calculate the spatially explicit RQ, we divide each chemical's internal exposure concentration, in vitro equivalent steady-state plasma concentration, by the hazard, $EC_{10}$ which is the concentration that produces that chemical's 10 % maximal response. This scales the exposure concentration by a standardized potency that is more conservative than the $EC_{50}$ used in the other examples and is typically the lowest concentrations that causes statistically significant cytotoxic effects in reporter gene assays (Escher et al., 2020b). The individual risk quotients are then summed to give an overall sum of risk quotients, given by Eq. (7):

$$RQ(s) = \Sigma_{i=1}^{n(s)} \frac{C_i(s)}{EC_{10,i}} \quad (7)$$

A sum of RQ < 1 indicates no risk of adverse effects, while a sum of RQ >1 indicates a risk of adverse effects. For simplicity, we will refer to this as RQ throughout the paper.

### 2.6. Monte Carlo uncertainty assessment

A Monte Carlo uncertainty assessment is used to quantify the parameter variability at different steps in the workflow. We explore three different sources of uncertainty; (1) in the air pollution estimates, we explore the variability in the county average, (2) in the toxicokinetics, we explore the interindividual variability of TK parameters, following the approach utilized in Ring et al. (2017), and (3) in the concentration-response model we explore the variability in model fit parameters. A detailed walk-through of the simulation of an individual and corresponding unit conversions is available in the Supplemental Information. In brief, we use the parameter distributions for the county-level contaminant concentration, age, obesity status, and concentration-response model parameters (top of the curve and $EC_{50}$) with a Monte Carlo framework to sample 1000 possible values using a truncated normal distribution. A binomial distribution was used for binary outcomes such as weight status (e.g., normal weight vs. obese). Once the model was created, the influences of various sources of uncertainty on the overall predictions of combined risk were determined by holding all parameters constant while the source being investigated was allowed to vary. The sources of uncertainty investigated include the uncertainty in the estimates of the fraction of obesity in the county populations and the age distributions, the concentration-response, the toxicokinetic parameters, and the estimate of the concentrations of chemicals in the air. The distribution of the risk metrics for all Monte Carlo simulations where only the specific parameter is allowed to vary while all other parameters are held constant using the median value is compared to the baseline where all parameters are allowed to vary simultaneously.

## 3. Results

Fig. 2 shows the spatial distribution, chemical co-occurrence, and relative potency as indicated by the individual chemical $EC_{50}$. The 41 chemicals are heterogeneously distributed

across space and range from ubiquitous occurrences, such as acetophenone and trifluralin, to hyper-local occurrences, such as 1,3-propane and heptachlor (Fig. 2A). The chemical co-occurrence (Fig. 2B) is also very heterogeneous, with regions of Colorado and Nebraska having the lowest number of co-occurring chemicals and the Texas and Louisiana Gulf region and the Ohio-Pennsylvania border having the highest number of co-occurring chemicals. These chemicals also range in potency for upregulating mRNA expression of CYP1A1 from an EC50 of 6.3 μM (ethylene glycol), which is the most potent, to 164.4 μM (2-chloroacetophenone), which is the least potent chemical. The median potency is 65.8 μM.

Fig. 3 shows the geospatial distribution of the predicted risk of increased mRNA expression of CYP1A1 over 1 year based on cumulative chemical exposures using CA (Fig. 3A), IA (Fig. 3B), and RQ (Fig. 3C). The minimum, median, and maximum values for each risk metric are summarized in Table 1. The results are broken down by percentiles generated from the Monte Carlo uncertainty assessment for quantifying interindividual variation in risk. The CA and IA risk metrics are approximately the same due to the low individual chemical concentrations. This is also observed in the histogram showing the distribution of all the risk metrics for 1000 Monte Carlo iterations (Fig. S1). At the 5th percentile, all risk metrics are small, and no counties have a RQ > 1. The median (50th percentile),) representing the average response for a county given the population variability shows patterns of elevated responses, and RQ >1 starts to emerge. While the median risk metrics are still in ranges that are not of concern, the maximum values are in a range of potential concern. At the 50th percentile, 2.6 % of counties have RQ > 1 (Fig. S2). Notably, patterns including central Colorado, northeastern Illinois, eastern Massachusetts, and the District of Columbia are emerging as higher than the surrounding areas. These patterns become more pronounced in the 95th percentile of risk showing additional areas of elevated risk, including southern California and Florida. The county median of the 95th percentile is approaching levels of concern, and the maximum value for predicted *Log2*-fold change in mRNA expression is 1.96. The maximum RQ, which used the $EC_{10}$, is 124.7 where 45.7 % of counties have an RQ > 1 (Fig. S2).

Results from the sensitivity analysis (Fig. 4) highlight how the different parameters used to generate the geospatial risk assessment affect the risk metrics. The predicted *log2* fold change in mRNA expression of CYP1A1 for CA (Fig. 4A) and IA (Fig. 4B) are influenced by all the tested parameters in the process, but the concentration-response parameters (response at the top of the curve and $EC_{50}$) and the external concentration exert the most variability. The population parameters of age, obesity status, and toxicokinetic parameters (also based on population parameters) have a smaller relative influence. The RQ (Fig. 4C) is most influenced by the concentration-response parameter and external concentration.

## 4. Discussion

In this paper, we presented a novel, geospatially informed risk assessment for complex mixtures that can be used to identify geographic regions that are potentially at an increased risk for molecular level perturbations in exposure-effect pathways. This method is complimentary but markedly different from the traditional apical endpoint-based approach (i.e., cancer and non-cancer) typically used in component-based cumulative risk assessments

and epidemiological studies, as our study uses response data for molecular targets of environmental chemicals that cause early events in AEP-AOP networks. This approach could be expanded to include additional assays that measure other early events in an AEP-AOP network to help understand the potential progression of an adverse outcome that results from combined exposure to multiple environmental chemicals. Here we present a proof-of-concept example using a perturbation of CYP1A1 activity. This perturbation could affect the activation of concurrent exposures to other chemicals and the risks such chemicals pose to individuals in certain counties. The approach could be applied to other early events, such as binding to receptors.

The mapped results highlight the heterogeneous nature of the predicted risk of molecular perturbation. While it is easier to contextualize the risk quotient, where any value over 1 is of concern, contextualizing the result of the $Log2$ fold change in CYP1A1 mRNA expression is more difficult. The transcription of the CYP1A1 is easily induced by smoking tobacco, where it can be induced up to 100-fold in the human lung (McLemore et al., 1990). Even at the highest risk scenario (95th percentile), the predicted maximum would be a 4-fold change. However, it is important to consider the time frame of exposure used in this study, which is 1 year (365 days). Since individuals are exposed to chemicals over a lifetime, the induction of this enzyme could be much higher. Our sensitivity analysis shows that these risk metrics are most affected by the external concentration and the uncertainty around the parameters that define the concentration-response curves. This provides further evidence that the geospatially explicit nature of exposure to environmental mixtures is important and should be considered in future mixtures research. However, it is important to recognize that this analysis does not consider the temporal variation in concentration patterns of different chemicals which may not vary similarly over time (i.e., diurnal differences with morning and evening peaks for traffic-related VOCs and daytime peaks for photoreactant secondary organics like formaldehyde, etc.).

The NATA only reports speciated data on a small fraction of an individual's possible exposome. In the cHTS data, 721 additional chemicals induce CYP1A1 transcripts in this assay, which includes pharmaceuticals (e.g., erythromycin and acetaminophen), pesticides (e.g., parathion and heptachlor), and other environmental contaminants (e.g., anthracene and phenanthrene). Further, Cytochrome P450 enzymes have been shown to have high interindividual variation caused by genetic polymorphisms, which vary by geographic ancestry (Vichi et al., 2021). Future work could integrate more representative information on population-level enzyme variability, including sex and geographic ancestry, to produce a more accurate estimation of internal contaminant concentrations and integrate additional assays for AhR, CYP1A1 enzyme activity, and other phase I metabolic enzymes.

The risk patterns observed in this study result from a unique combination of external concentration, chemical co-occurrence, and chemical characteristics such as potency that drive the metabolic clearance ($C_{SS}$ Fig. S3) highlighting that co-occurrence alone may not be a good predictor of risk. Metabolic clearance and the spatial factors (age and obesity; Fig. S4 and Fig. S5, respectively) that drive these patterns influence the spatial distribution of the risk metrics. For example, our results show that the Texas gulf region has a high number of co-occurring chemicals but did not have an elevated risk of CYP1A1 perturbation

compared to other regions of the USA. Comparatively, the Chicago area had a moderate number of co-occurring chemicals, yet this same region was among the highest predicted risk of increasing CYP1A1 mRNA expression. The geographic variability observed in the Texas gulf region may be due to lower concentrations of chemicals and younger average population age compared to Chicago and thus a higher population-level metabolic clearance of the potent chemicals.

It is known that mixture modeling with real-world exposures is complicated and understudied, especially for the low concentration typically observed in the environment (Escher et al., 2020a). We explore three different methods for quantifying the risk of molecular perturbation. Interestingly, the CA and IA methods results are nearly identical, which has also been observed in samples of ambient environmental mixtures (Escher et al., 2018; Neale et al., 2020). Escher et al. (2020a) demonstrate that CA/IA can be used jointly with high certainty for predicting the $EC_{10}$ and that it is pragmatic to use a joint CA/IA model up to an $EC_{30}$. Even in the highest exposure scenario (95th percentile), the ambient environmental air concentrations were low and never left the low-dose linear region of the concentration-response curve where CA and IA were approximately equal. CA and IA may deviate in other applications where concentrations are high, such as near known point sources.

A primary objective of NATA is also risk characterization, in which the estimated external concentrations are integrated with concentration-response models to quantify potential risk for developing cancer and non-cancer endpoints. Several key differences between our results and the NATA health effects make it difficult to make a direct comparison. First, is the point along the biological pathways that our method informs on; the cHTS assays inform on key molecular events early in disease progression. In contrast, the NATA health risk assessment relies on animal model in vivo data from the Integrated Risk Information System (IRIS) to inform on an apical health endpoint. Many biological steps must occur to progress from a molecular perturbation to an adverse health outcome. Our bottom-up approach does not yet consider the biological complexity that results in an adverse outcome; however, our method is useful for early warning by identifying regions for which perturbation of a particular molecular target may contribute to an adverse outcome. Second, the timeframes are different between the two risk assessments, we use a 1-year risk based on the 2014 chemical averages in air, and NATA uses a lifetime (70 years) risk. For these reasons and the assay chosen for this proof-of-concept, the two risk assessments are complimentary, but they are not directly comparable.

The CYP1A1 enzyme is involved in a variety of biological disease pathways, which our proof-of-concept assay informs on by measuring the mRNA expression of the enzyme. Most notable is the role that the metabolism of chemicals by the CYP1A1 enzyme plays in the metabolic activation of procarcinogens that form DNA adducts and reactive oxygen species, which can lead to cancer. Due to this, CYP1A1 has been implicated in cancer pathways, including lung, colon, liver, breast, and prostate, and polymorphisms in the CYP1A1 gene can increase cancer risk (Androutsopoulos et al., 2009; Nebert and Dalton, 2006; Uhlen et al., 2017). CYP1A1 induction is also a non-specific biomarker for AhR activation (Hu et al., 2007). When activated, AhR can interact with nuclear factors and receptors such

as oestrogen receptor-α (ESRα) and retinoblastoma protein 1 (RB1), which can result in abnormal growth leading to the development of cancer (Nebert and Dalton, 2006). This assay measures a target that could be incorporated into an AOP, and as AOP development expands, more geospatially explicit chemical mixtures can be causally associated with apical health outcomes. An important next step is to use this proposed workflow to link molecular perturbations to adverse outcomes. Future work should focus on diseases with well-defined etiologies and established AOPs.

This workflow relies on various input parameters and assumptions, which can impact the results. For example, in our proof-of-concept example, we assumed that 100 % of the inhaled chemical was converted into an internal concentration for simplicity. If the information about bioavailability or the absorbed fraction of the chemicals is known, this could be easily incorporated into the equation that estimates internal concentration from external concentrations. We also assume that the cumulative internal dose over one-year results from repeated exposures that maintain the steady-state plasma concentration. Together these assumptions are likely an overestimation of exposure and risk. Still, this approach is protective of human health as per the “worse-case-scenario” risk assessment method (OECD, 2018). These limitations provide opportunities for future research that can more accurately estimate the internal concentrations of chemicals through advancing NAMs related to cumulative chemical exposures and properties of absorption, distribution, metabolism, and excretion (ADME).

We also chose to use a 2-parameter hill model, with a fixed slope of one, rather than a 3-parameter hill model with a variable slope. Since our probabilistic approach considers the full uncertainty of each concentration-response parameter (given by the standard deviation), the maximal efficacy of individual concentration-response relationships is not always below the estimated mixture effect (Figs. S6 and S7). This prevents the use of concentration-response curves with three or more parameters since the inverse function will not be defined which requires a more complex GCA model (Scholze et al., 2014). However, using different concentration-response curves may impact the results. If the slope was originally close to one in the 3-parameter model, then there is little difference in the $EC_{50}$ and top-of-the-curve values. In our example, in the 3-parameter hill models, the minimum $EC_{50}$ is 9.0 μM (pentachlorophenol), the median $EC_{50}$ is 46.6 μM, and the maximum $EC_{50}$ is 87.2 μM (1,6-diisocyanatohexane), whereas in the 2-parameter model, the minimum $EC_{50}$ values are similar (minimum = 6.2 μM), the median value is about 20 μM higher (median = 65.8 μM), and the maximum value is almost double (maximum = 164.4 μM). Thus, using the 2-parameter hill model may underestimating risk. The modular nature of this workflow allows individuals to select input parameters that are best suited for their research question. This workflow can be fit-for-purpose as outlined in Table 2.

In conclusion, we have presented a novel method that integrated geospatial environmental exposure estimates with chemical hazard data from the Tox21 cHTS assays. This provides an alternative method of conducting risk assessment, taking a bottom-up approach where complex chemical exposures are integrated via common molecular targets instead of apical endpoints traditionally used in risk assessment. This approach can provide an early warning based on localized chemical exposure data that may contribute to adverse outcomes

where the molecular target is implicated. This method will help support the development of models to better predict how chemicals affect biological responses. As such, this geospatial workflow can also support the development of AOPs as our approach aids in hypothesis generation related to the potential effects of exposure to environmentally relevant chemicals and the molecular targets that they may impact. Elucidating and strengthening causal relationships between non-apical and apical endpoints will be key in helping to further develop this proof-of-concept as a method for the early identification of potential adverse outcomes. This fit-for-purpose workflow could be expanded to other geocoded environmental samples such as water, assays/molecular targets, chemicals, and risk metrics. Future work that will aid in the utility of this workflow also includes improving and expanding chemical exposure assessments in air and other media.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.158905.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by the Intramural Research Program of National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH), ZIA ES103316-04, ZIA ES103373-01, Spatiotemporal Health Analytics project ZIA ES103368-02, and Contract No. HHSN273201500010C. The authors would like to thank Drs. David Hines, Bonnie Joubert, Havala Pye, and Rogelio Tornero-Velez and for their helpful U.S. EPA and Division of the National Translational Toxicology internal reviews of the manuscript. Note the Division of Translational Toxicology is formerly the Division of the National Toxicology Program.

## Data availability

Data will be made available on request.

## References

Abedini J, Cook B, Bell S, Chang X, Choksi N, Daniel AB, Hines D, Karmaus AL, Mansouri K, McAfee E, Phillips J, Rooney J, Sprankle C, Allen D, Casey W, Kleinstreuer N, 2021. Application of new approach methodologies: ICE tools to support chemical evaluations. Comput. Toxicol 20, 100184. 10.1016/j.comtox.2021.100184.

Altenburger R, Backhaus T, Boedeker W, Faust M, Scholze M, Grimme LH, 2000. Predictability of the toxicity of multiple chemical mixtures to Vibrio fischeri: mixtures composed of similarly acting chemicals. Environ. Toxicol. Chem 19, 2341. 10.1897/1551-5028(2000)019<2341:pottom>2.3.co;2.

Androutsopoulos VP, Tsatsakis AM, Spandidos DA, 2009. Cytochrome P450 CYP1A1: wider roles in cancer progression and prevention. BMC Cancer 9, 1–17. 10.1186/1471-2407-9-187. [PubMed: 19118499]

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL, 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem 29, 730–741. 10.1002/etc.34. [PubMed: 20821501]

Appel KW, Bash JO, Fahey KM, Foley KM, Gilliam RC, Hogrefe C, Hutzell WT, Kang D, Mathur R, Murphy BN, Napelenok SL, Nolte CG, Pleim JE, Pouliot GA, Pye HOT, Ran L, Roselle SJ, Sarwar G, Schwede DB, Sidi FI, Spero TL, Wong DC, 2021. The Community Multiscale Air Quality

(CMAQ) model versions 5.3 and 5.3.1: system updates and evaluation. Geosci. Model Dev 14, 2867–2897. 10.5194/gmd-14-2867-2021. [PubMed: 34676058]

Auerbach S, Filer D, Reif D, Walker V, Holloway AC, Schlezinger J, Srinivasan S, Svoboda D, Judson R, Bucher JR, Thayer KA, 2016. Prioritizing environmental chemicals for obesity and diabetes outcomes research: a screening approach using ToxCastTM high-throughput data. Environ. Health Perspect 124, 1141–1154. 10.1289/ehp.1510456. [PubMed: 26978842]

Backhaus T, Altenburger R, Boedeker W, Faust M, Scholze M, Grimme LH, 2000. Predictability of the toxicity of a multiple mixture of dissimilarly acting chemicals to Vibrio fischeri. Environ. Toxicol. Chem 19, 2348–2356. 10.1002/etc.5620190927.

Berenbaum MC, 1985. The expected effect of a combination of agents: the general solution. J. Theor. Biol 114, 413–431. 10.1016/S0022-5193(85)80176-4. [PubMed: 4021503]

Breen M, Ring CL, Kreutz A, Goldsmith MR, Wambaugh JF, 2021. High-throughput PBTK models for in vitro to in vivo extrapolation. Expert Opin. Drug Metab. Toxicol 00, 1–19. 10.1080/17425255.2021.1935867.

Centers for Disease Control and Prevention, 2021. PLACES [WWW Document]. URL https://www.cdc.gov/places (accessed 10.16.21).

Cleland SE, West JJ, Jia Y, Reid S, Raffuse S, O'Neill S, Serre ML, 2020. Estimating wildfire smoke concentrations during the October 2017 California fires through BME space/time data fusion of observed, modeled, and satellite-derived PM2.5. Environ. Sci. Technol 54, 13439–13447. 10.1021/acs.est.0c03761. [PubMed: 33064454]

Dennis KK, Marder E, Balshaw DM, Cui Y, Lynes MA, Patti GJ, Rappaport SM, Shaughnessy DT, Vrijheid M, Barr DB, 2017. Biomonitoring in the era of the exposome. Environ. Health Perspect 125, 502–510. 10.1289/EHP474. [PubMed: 27385067]

Escher BI, Neale PA, Villeneuve DL, 2018. The advantages of linear concentration–response curves for in vitro bioassays with environmental samples. Environ. Toxicol. Chem 37, 2273–2280. 10.1002/etc.4178. [PubMed: 29846006]

Escher BI, Braun G, Zarfl C, 2020a. Exploring the concepts of concentration addition and independent action using a linear low-effect mixture model. Environ. Toxicol. Chem 39, 2552–2559. 10.1002/etc.4868. [PubMed: 32897547]

Escher BI, Henneberger L, König M, Schlichting R, Fischer FC, 2020b. Cytotoxicity burst? Differentiating specific from nonspecific effects in tox21 in vitro reporter gene assays. Environ. Health Perspect 128, 1–10. 10.1289/EHP6664.

Escher BI, Stapleton HM, Schymanski EL, 2020c. Tracking complex mixtures of chemicals in our changing environment. Science 367, 388–392. 10.1126/science.aay6636.Tracking. [PubMed: 31974244]

Fay KA, Villeneuve DL, Swintek J, Edwards SW, Nelms MD, Blackwell BR, Ankley GT, 2018. Differentiating pathway-specific from nonspecific effects in high-throughput toxicity data: a foundation for prioritizing adverse outcome pathway development. Toxicol. Sci 163, 500–515. 10.1093/toxsci/kfy049. [PubMed: 29529260]

Filer DL, Kothiya P, Woodrow Setzer R, Judson RS, Martin MT, 2017. Tcpl: the ToxCast pipeline for high-throughput screening data. Bioinformatics 33, 618–620. 10.1093/bioinformatics/btw680. [PubMed: 27797781]

Franzosa JA, Bonzo JA, Jack J, Baker NC, Kothiya P, Witek RP, Hurban P, Siferd S, Hester S, Shah I, Ferguson SS, Houck KA, Wambaugh JF, 2021. High-throughput toxicogenomic screening of chemicals in the environment using metabolically competent hepatic cell cultures. npj Syst.Biol. Appl 7. 10.1038/s41540-020-00166-2.

Hines DE, Edwards SW, Conolly RB, Jarabek AM, 2018. A case study application of the aggregate exposure pathway (AEP) and adverse outcome pathway (AOP) frameworks to facilitate the integration of human health and ecological end points for cumulative risk assessment (CRA). Environ. Sci. Technol 52, 839–849. 10.1021/acs.est.7b04940. [PubMed: 29236470]

Hines DE, Conolly RB, Jarabek AM, 2019. 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. Environ. Sci. Technol 53, 11002–11012. 10.1021/acs.est.9b04639. [PubMed: 31436975]

Howard GJ, Schlezinger JJ, Hahn ME, Webster TF, 2010. Generalized concentration addition predicts joint effects of aryl hydrocarbon receptor agonists with partial agonists and competitive antagonists. Environ. Health Perspect 118, 666–672. 10.1289/ehp.0901312. [PubMed: 20435555]

Hsieh NH, Chen Z, Rusyn I, Chiu WA, 2021. Risk characterization and probabilistic concentration–response modeling of complex environmental mixtures using new approach methodologies (Nams) data from organotypic in vitro human stem cell assays. Environ. Health Perspect 129, 1–13. 10.1289/EHP7600.

Hu W, Sorrentino C, Denison MS, Kolaja K, Fielden MR, 2007. Induction of Cyp1a1 is a nonspecific biomarker of aryl hydrocarbon receptor activation: results of large scale screening of pharmaceuticals and toxicants in vivo and in vitro. Mol. Pharmacol 71, 1475–1486. 10.1124/mol.106.032748. [PubMed: 17327465]

Huang K, Sanchez SA, Eckelman MJ, 2021. Using the US National air Toxics Assessment to benchmark the USEtox inhalation - mediated carcinogenic impacts of air emissions. Int. J. Life Cycle Assess 1417–1430. 10.1007/s11367-021-01918-w.

Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, Dellarco V, Henry T, Holderman T, Sayre P, Tan S, Carpenter T, Smith E, 2009. The toxicity data landscape for environmental chemicals. Environ. Health Perspect 117, 685–695. 10.1289/ehp.0800168. [PubMed: 19479008]

Judson RS, Kavlock RJ, Setzer RW, Cohen Hubal EA, Martin MT, Knudsen TB, Houck KA, Thomas RS, Wetmore BA, Dix DJ, 2011. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. Chem. Res. Toxicol 24, 451–462. 10.1021/tx100428e. [PubMed: 21384849]

Judson RS, Magpantay FM, Chickarmane V, Haskell C, Tania N, Taylor J, Xia M, Huang R, Rotroff DM, Filer DL, Houck KA, Martin MT, Sipes N, Richard AM, Mansouri K, Woodrow Setzer R, Knudsen TB, Crofton KM, Thomas RS, 2015. Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. Toxicol. Sci 148, 137–154. 10.1093/toxsci/kfv168. [PubMed: 26272952]

Katzfuss M, 2017. A multi-resolution approximation for massive spatial datasets. J. Am. Stat. Assoc 112, 201–214. 10.1080/01621459.2015.1123632.

Katzfuss M, Guinness J, Gong W, Zilber D, 2020. Vecchia approximations of Gaussian-process predictions. J. Agric. Biol. Environ. Stat 25, 383–414. 10.1007/s13253-020-00401-7.

Kavlock RJ, Bahadori T, Barton-Maclaren TS, Gwinn MR, Rasenberg M, Thomas RS, 2018. Accelerating the pace of chemical risk assessment. Chem. Res. Toxicol 31, 287–290. 10.1021/acs.chemrestox.7b00339. [PubMed: 29600706]

Kleinstreuer NC, Sullivan K, Allen D, Edwards S, Mendrick DL, Embry M, Matheson J, Rowlands JC, Munn S, Maull E, Casey W, 2016. Adverse outcome pathways: from research to regulation scientific workshop report. Regul. Toxicol. Pharmacol 76, 39–50. 10.1016/j.yrtph.2016.01.007. [PubMed: 26774756]

Kleinstreuer NC, Ceger P, Watt ED, Martin M, Houck K, Browne P, Thomas RS, Casey WM, Dix DJ, Allen D, Sakamuru S, Xia M, Huang R, Judson R, 2017. Development and validation of a computational model for androgen receptor activity. Chem. Res. Toxicol 30, 946–964. 10.1021/acs.chemrestox.6b00347. [PubMed: 27933809]

Kortenkamp A, Faust M, 2018. Regulate to reduce chemical mixture risk. Science (80-.) 361, 224–226. 10.1126/science.aat9219.

Kortenkamp A, Faust M, Scholze M, Backhaus T, 2007. Low-level exposure to multiple chemicals: reason for human health concerns? Environ. Health Perspect 115, 106–114. 10.1289/ehp.9358. [PubMed: 18174958]

Krewski D, Andersen ME, Mantus E, Zeise L, Kavlock RJ, Austin CP, Tice RR, 2009. Toxicity testing in the 21st century: implications for human health risk assessment. Risk Anal. 29, 485–487. 10.1111/j.1539-6924.2008.01168.x. [PubMed: 19076321]

McLemore TL, Adelberg S, Liu MC, Mcmahon NA, Yu SJ, Hubbard WC, Czerwinski M, Wood TG, Storeng R, Lubet RA, Eggleston JC, Boyd MR, Hines RN, 1990. Expression of CYP1A1 gene in patients with lung cancer: evidence for cigarette smoke-induced gene expression in normal lung tissue and for altered gene regulation in primary pulmonary carcinomas. J. Natl. Cancer Inst 82, 1333–1339. 10.1093/jnci/82.16.1333. [PubMed: 2380990]

Mescher M, Haarmann-Stemmann T, 2018. Modulation of CYP1A1 metabolism: from adverse health effects to chemoprevention and therapeutic options. Pharmacol. Ther 187, 71–87. 10.1016/j.pharmthera.2018.02.012. [PubMed: 29458109]

Messier KP, Katzfuss M, 2021. Scalable penalized spatiotemporal land-use regression for ground-level nitrogen dioxide. Ann. Appl. Stat 15, 688–710. 10.1214/20-AOAS1422. [PubMed: 35069963]

Messier KP, Kane E, Bolich R, Serre ML, 2014. Nitrate variability in groundwater of North Carolina using monitoring and private well data models. Environ. Sci. Technol 48, 10804–10812. 10.1021/es502725f. [PubMed: 25148521]

Miller GW, Jones DP, 2014. The nature of nurture: refining the definition of the exposome. Toxicol. Sci 137, 1–2. 10.1093/toxsci/kft251. [PubMed: 24213143]

National Research Council, 1983. Risk Assessment in the Federal Government: Managing the Process. 10.17226/317 Washington (DC).

Neale PA, Braun G, Brack W, Carmona E, Gunold R, König M, Krauss M, Liebmann L, Liess M, Link M, Schäfer RB, Schlichting R, Schreiner VC, Schulze T, Vormeier P, Weisner O, Escher BI, 2020. Assessing the mixture effects in in vitro bioassays of chemicals occurring in small agricultural streams during rain events. Environ. Sci. Technol 54, 8280–8290. 10.1021/acs.est.0c02235. [PubMed: 32501680]

Nebert DW, Dalton TP, 2006. The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. Nat. Rev. Cancer 6, 947–960. 10.1038/nrc2015. [PubMed: 17128211]

OECD, 2018. Considerations for Assessing the Risks of Combined Exposure to Multiple Chemicals, Series on Testing And Assessment No. 296. Environment, Health and Safety Division Environment Directorate.

Pearce RG, Setzer RW, Strope CL, Wambaugh JF, Sipes NS, 2017a. httk: R package for high-throughput toxicokinetics. J. Stat. Softw 79, 1–26. 10.18637/jss.v079.i04.Submit. [PubMed: 30220889]

Pearce R, Setzer RW, Davis J, Wambaugh JF, 2017b. Evaluation and calibration of high-throughput predictions of chemical distribution to tissues. J. Pharmacokinet. Pharmacodyn 44, 549–565. 10.1007/s10928-017-9548-7.Evaluation. [PubMed: 29032447]

Price PS, Jarabek AM, Burgoon LD, 2020. Organizing mechanism-related information on chemical interactions using a framework based on the aggregate exposure and adverse outcome pathways. Environ. Int 138. 10.1016/j.envint.2020.105673.

Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, Yang C, Rathman J, Martin MT, Wambaugh JF, Knudsen TB, Kancherla J, Mansouri K, Patlewicz G, Williams AJ, Little SB, Crofton KM, Thomas RS, 2016. ToxCast chemical landscape: paving the road to 21st century toxicology. Chem. Res. Toxicol 29, 1225–1251. 10.1021/acs.chemrestox.6b00135. [PubMed: 27367298]

Ring CL, Pearce RG, Setzer RW, Wetmore BA, Wambaugh JF, 2017. Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. Environ. Int 106, 105–118. 10.1016/j.envint.2017.06.004. [PubMed: 28628784]

Ring CL, Arnot JA, Bennett DH, Egeghy PP, Fantke P, Huang L, Isaacs KK, Jolliet O, Phillips KA, Price PS, Shin HM, Westgate JN, Setzer RW, Wambaugh JF, 2019. Consensus modeling of median chemical intake for the U.S. population based on predictions of exposure pathways. Environ. Sci. Technol 53, 719–732. 10.1021/acs.est.8b04056. [PubMed: 30516957]

Scholz S, Nichols JW, Escher BI, Ankley GT, Altenburger R, Blackwell B, Brack W, Burkhard L, Collette TW, Doering JA, Ekman D, Fay K, Fischer F, Hackermüller J, Hoffman JC, Lai C, Leuthold D, Martinovic-Weigelt D, Reemtsma T, Pollesch N, Schroeder A, Schüürmann G, von Bergen M, 2022. The eco-exposome concept: supporting an integrated assessment of mixtures of environmental chemicals. Environ. Toxicol. Chem 41, 30–45. 10.1002/etc.5242. [PubMed: 34714945]

Scholze M, Silva E, Kortenkamp A, 2014. Extending the applicability of the dose addition model to the assessment of chemical mixtures of partial agonists by using a novel toxic unit extrapolation method. PLoS One 9. 10.1371/journal.pone.0088808.

Sipes NS, Wambaugh JF, Pearce R, Auerbach SS, Wetmore BA, Hsieh JH, Shapiro AJ, Svoboda D, Devito MJ, Ferguson SS, 2017. An intuitive approach for predicting potential human health risk with the Tox21 10k library. Environ. Sci. Technol 51, 10786–10796. 10.1021/acs.est.7b00650. [PubMed: 28809115]

Tan YM, Leonard JA, Edwards S, Teeguarden J, Egeghy P, 2018. Refining the aggregate exposure pathway. Environ. Sci. Process. Impacts 20, 428–436. 10.1039/c8em00018b. [PubMed: 29465734]

Teeguarden JG, Tan YM, Edwards SW, Leonard JA, Anderson KA, Corley RA, Kile ML, Simonich SM, Stone D, Tanguay RL, Waters KM, Harper SL, Williams DE, 2016. Completing the link between exposure science and toxicology for improved environmental health decision making: the aggregate exposure pathway framework. Environ. Sci. Technol 50, 4579–4586. 10.1021/acs.est.5b05311. [PubMed: 26759916]

U.S. EPA, 1992. Guidelines for Exposure Assessment Washington (DC).

U.S. EPA, 2011. Exposure Factors Handbook 2011 Edition (Final Report) Washington (DC).

U.S. EPA, 2018. Technical Support Document EPA's 2014 National Air Toxics Assessment. Research Triangle Park (NC).

Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, Sanli K, Von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J, Schwenk JM, Brunnström H, Glimelius B, Sjöblom T, Edqvist PH, Djureinovic D, Micke P, Lindskog C, Mardinoglu A, Ponten F, 2017. A pathology atlas of the human cancer transcriptome. Science (80-.) 357. 10.1126/science.aan2507.

Vermeulen R, Schymanski EL, Barabási AL, Miller GW, 2020. The exposome and health: where chemistry meets biology. Science 367, 392–396. 10.1126/science.aay3164. [PubMed: 31974245]

Vichi S, Buratti FM, Di Consiglio E, Turco L, Lautz LS, Darney K, Dorne JLCM, Testai E, 2021. OpenCYP: an open source database exploring human variability in activities and frequencies of polymophisms for major cytochrome P-450 isoforms across world populations. Toxicol. Lett 350, 267–282. 10.1016/j.toxlet.2021.07.019. [PubMed: 34352333]

Villeneuve DL, Angrish MM, Fortin MC, Katsiadaki I, Leonard M, Margiotta-Casaluci L, Munn S, O'Brien JM, Pollesch NL, Smith LC, Zhang X, Knapen D, 2018. Adverse outcome pathway networks II: network analytics. Environ. Toxicol. Chem 37, 1734–1748. 10.1002/etc.4124. [PubMed: 29492998]

Wambaugh JF, Wang A, Dionisio KL, Frame A, Egeghy P, Judson R, Setzer RW, 2014. High throughput heuristics for prioritizing human exposure to environmental chemicals. Environ. Sci. Technol 48, 12760–12767. 10.1021/es503583j. [PubMed: 25343693]

Wambaugh JF, Bare JC, Carignan CC, Dionisio KL, Dodson RE, Jolliet O, Liu X, Meyer DE, Newton SR, Phillips KA, Price PS, Ring CL, Shin HM, Sobus JR, Tal T, Ulrich EM, Vallero DA, Wetmore BA, Isaacs KK, 2019. New approach methodologies for exposure science. Curr. Opin. Toxicol 15, 76–92. 10.1016/j.cotox.2019.07.001.

Wang M, Sampson PD, Hu J, Kleeman M, Keller JP, Olives C, Szpiro AA, Vedal S, Kaufman JD, 2016. Combining land-use regression and chemical transport modeling in a spatiotemporal geostatistical model for ozone and PM2.5. Environ. Sci. Technol 50, 5111–5118. 10.1021/acs.est.5b06001. [PubMed: 27074524]

Watt J, Webster TF, Schlezinger JJ, 2016. Generalized concentration addition modeling predicts mixture effects of environmental PPAR$\gamma$ agonists. Toxicol. Sci 153, 18–27. 10.1093/toxsci/kfw100. [PubMed: 27255385]

Wild CP, 2005. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol. Biomark. Prev 14, 1847–1850. 10.1158/1055-9965.EPI-05-0456.

Zavala J, Freedman AN, Szilagyi JT, Jaspers I, Wambaugh JF, Higuchi M, Rager JE, 2020. New approach methods to evaluate health risks of air pollutants: critical design considerations for in vitro exposure testing.

Zhu H, Chinthakindi S, Kannan K, 2021. A method for the analysis of 121 multi-class environmental chemicals in urine by high-performance liquid chromatography-tandem mass spectrometry. J. Chromatogr. A 1646, 462146. 10.1016/j.chroma.2021.462146. [PubMed: 33895641]

## HIGHLIGHTS

- We assess the geographic variation for the joint effect of many chemical exposures.
- This example workflow integrates NAMs with chemical exposure data.
- The biological perturbations were heterogeneously distributed across space.
- Exposure concentrations, demographics, and toxicokinetics influence variability.
- We provide methods for modeling the source-exposure-effect continuum.

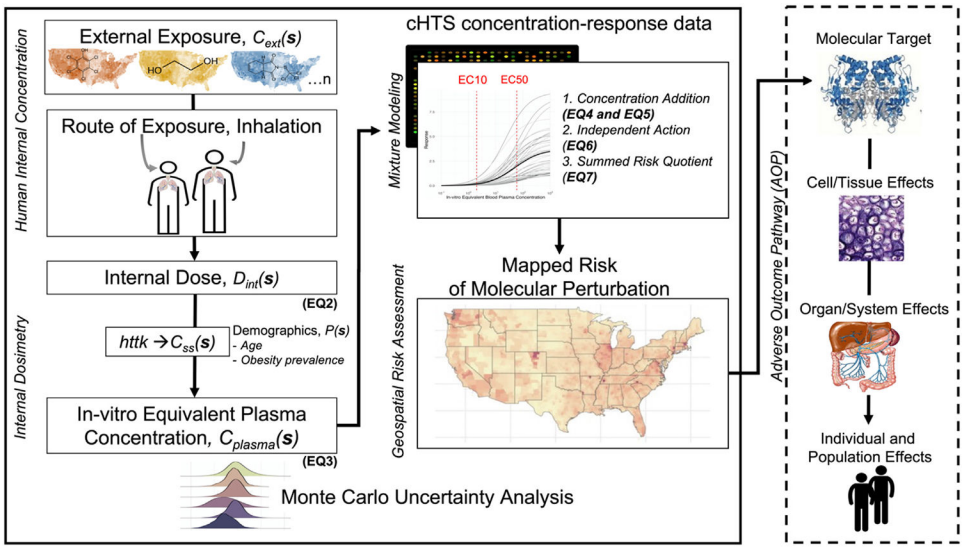



**Fig. 1.**

Schematic overview of the methods used to quantify the geospatial risk of molecular perturbation. The spatially explicit external chemical concentrations ($C_{ext}(s)$) are converted to an internal dose ($D_{int}(s)$) via inhalation parameters and using county-specific distributions of age and obesity prevalence ($P(s)$). $D_{int}(s)$ is then converted into an in vitro equivalent steady-state plasma concentration ($C_{plasma}(s)$) using a steady-state plasma conversion factor ($C_{ss}(s)$) estimated using physiological based toxicokinetic (PBTK) modeling. The dashed box indicates that $P(s)$ influences both the inhalation rate and $C_{ss}(s)$ used to calculate $C_{plasma}(s)$. The geospatial risk assessment $R(s)$ is developed by using the $C_{plasma}(s)$ and concentration-response models generated using curated high-throughput (cHTS) data, which are integrated using mixtures models (e.g., concentration addition) to predict the risk of exposure. We use a probabilistic framework using a Monte Carlo uncertainty analysis. The mapped risk of molecular perturbation can be integrated with the Adverse Outcome Pathway (AOP) framework to link the molecular-level perturbation to an adverse outcome at the individual and population levels.

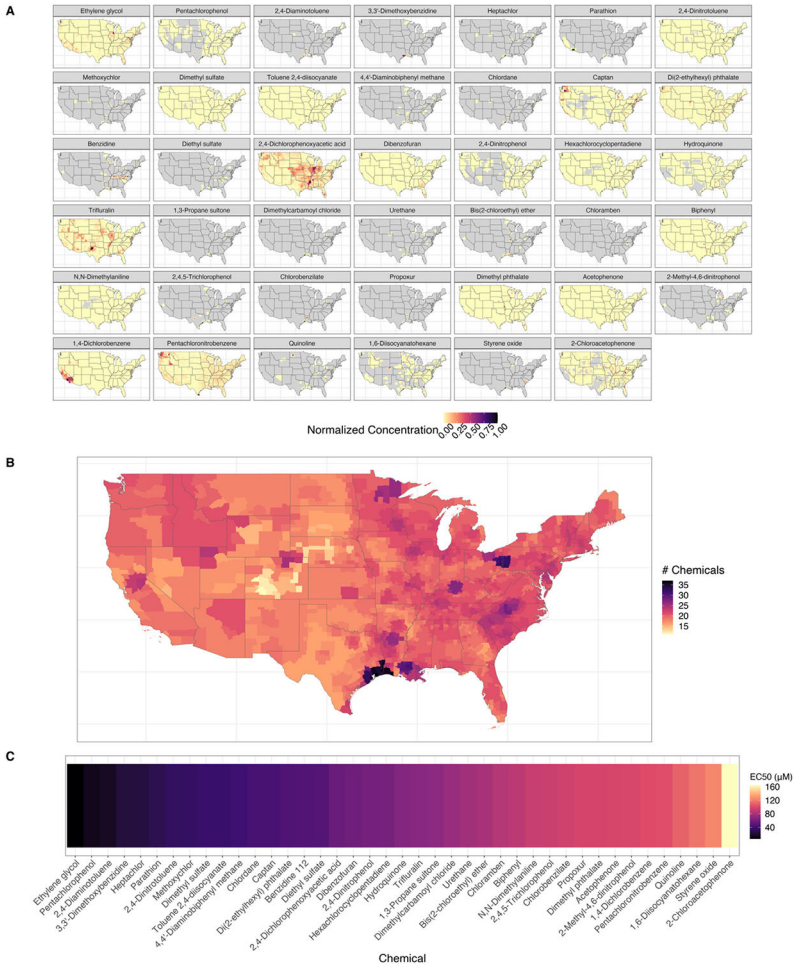



**Fig. 2.**

(A) Map of contaminant distributions for individual chemicals. Values are range scaled between zero and one to make chemicals of different magnitudes comparable. Black indicates higher environmental chemical concentrations in the air, and grey indicates counties where the chemical does not occur. The chemicals are ordered by the $EC_{50}$. (B) The number of chemicals that co-occur in each county results in geospatially unique chemical mixtures. Here black indicates a higher number of chemicals present in the county that have positive hit calls for the CYP1A1 mRNA transcription assay. (C) The $EC_{50}$ (the effective concentration at 50 % activity) for all chemicals is based on the refit 2-parameter Hill model. The black colour indicates lower $EC_{50}$ values, which indicates a more potent chemical.

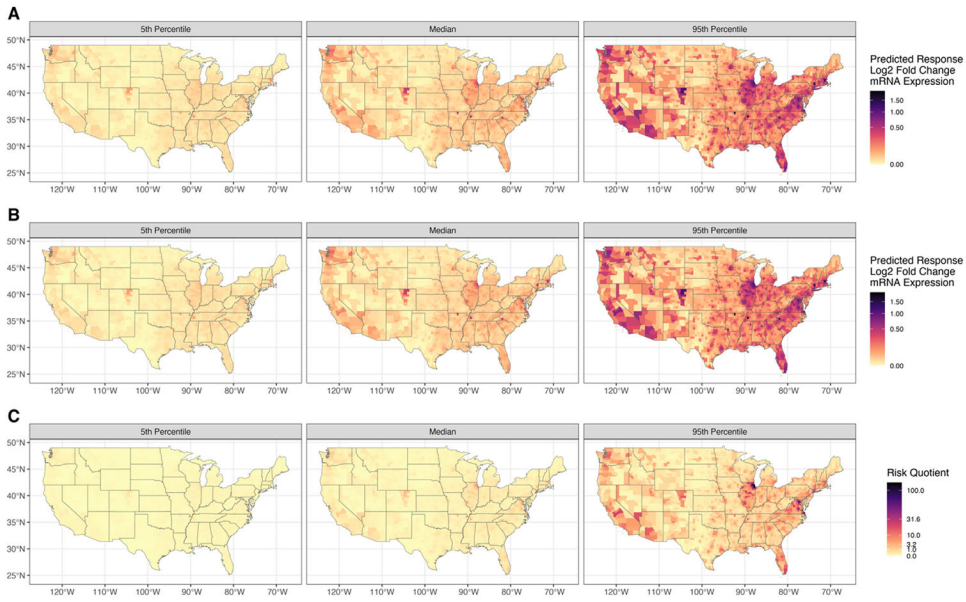



**Fig. 3.**

The predicted 5th, 50th (median), and 95th percentiles of predicted risk of increased mRNA expression of CYP1A1 using (A) Concentration Addition, (B) Independent Action, and (C) Risk Quotient using the 10 % effective concentration ($EC_{10}$). In these figures, yellow indicates lower risk, and black indicates higher risk.

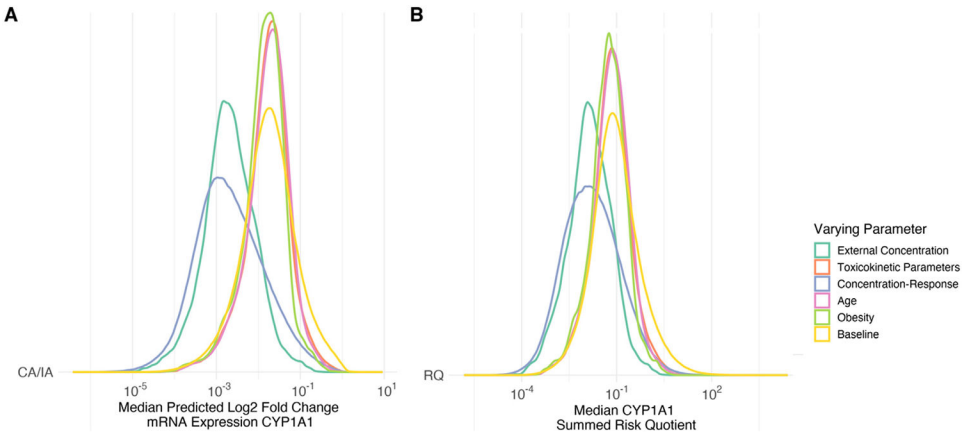



**Fig. 4.**

Probability density functions of the predicted risk of increased mRNA expression of CYP1A1 using (A) Concentration Addition (CA)/Independent Action (IA) and (B) The sum of the Risk Quotients (RQ) using the 10 % effective concentration ($EC_{10}$) under different Monte Carlo uncertainty assessment. The varying parameter is allowed to vary while all other parameters are held constant using the median value. The yellow histogram is the baseline where all parameters are allowed to vary.

## Table 1

Summary statistics for the minimum, median, and maximum values of risk metric results from the Monte Carlo uncertainty assessment in the 5th percentile, median (50th percentile), and 95th percentile for concentration addition (CA), independent action (IA), where the unit for both is the Predicted Response (Log2 fold change in mRNA expression) and the risk quotient (RQ) which is unitless.

| Risk metric | 5th Percentile | | | Median (50th Percentile) | | | 95th Percentile | | |
|---|---|---|---|---|---|---|---|---|---|
| | CA | IA | RQ | CA | IA | RQ | CA | IA | RQ |
| Minimum | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Median | 0.00 | 0.00 | 0.02 | 0.02 | 0.02 | 0.09 | 0.11 | 0.11 | 0.87 |
| Maximum | 0.12 | 0.12 | 0.46 | 0.73 | 0.74 | 5.75 | 1.96 | 1.97 | 124.70 |

**Table 2**

Summary of flexibilities in different workflow steps to design a fit-for-purpose geospatial risk assessment of exposure to environmental chemical mixtures via a common molecular target.

| Workflow steps | Possible variations | Details |
|---|---|---|
| Geographic extent | Local, Regional, National | • Selected based on available geospatial data and region of interest |
| Molecular target | Assay system | • Selected based on the molecular target, system, or adverse outcome pathways of interest<br>• Selection constrained based on overlap with relevant and available external exposure data<br>• Alternatively, this can be selected to match an apical outcome of interest |
| External exposure | Geospatial model<br>Environmental measurements | • Selected based on available geospatial data |
| Route of exposure/ internal concentration | Population parameters<br>Route of exposure<br>Internal exposure calculation<br>Human biomonitoring data | • The route of exposure (e.g., oral, dermal, inhalation) and population (e.g., adults vs. children) of interest can be selected, and internal concentrations can be calculated based on equations specific to that population using the US EPA Exposure Factors Handbook (U.S. EPA, 2011)<br>• Exposure factors like absorption can be taken into consideration at this step<br>• Alternatively, geotagged biomonitoring data can be used instead of estimates |