

Using *In Vitro* High-Throughput Screening Data for Predicting Benzo[k]Fluoranthene Human Health Hazards

Lyle D. Burgoon,^{1,*} Ingrid L. Druwe,² Kyle Painter,² and Erin E. Yost²

Today there are more than 80,000 chemicals in commerce and the environment. The potential human health risks are unknown for the vast majority of these chemicals as they lack human health risk assessments, toxicity reference values, and risk screening values. We aim to use computational toxicology and quantitative high-throughput screening (qHTS) technologies to fill these data gaps, and begin to prioritize these chemicals for additional assessment. In this pilot, we demonstrate how we were able to identify that benzo[k]fluoranthene may induce DNA damage and steatosis using qHTS data and two separate adverse outcome pathways (AOPs). We also demonstrate how bootstrap natural spline-based meta-regression can be used to integrate data across multiple assay replicates to generate a concentration-response curve. We used this analysis to calculate an *in vitro* point of departure of 0.751 μM and risk-specific *in vitro* concentrations of 0.29 μM and 0.28 μM for 1:1,000 and 1:10,000 risk, respectively, for DNA damage. Based on the available evidence, and considering that only a single HSD17B4 assay is available, we have low overall confidence in the steatosis hazard identification. This case study suggests that coupling qHTS assays with AOPs and ontologies will facilitate hazard identification. Combining this with quantitative evidence integration methods, such as bootstrap meta-regression, may allow risk assessors to identify points of departure and risk-specific internal/*in vitro* concentrations. These results are sufficient to prioritize the chemicals; however, in the longer term we will need to estimate external doses for risk screening purposes, such as through margin of exposure methods.

KEY WORDS: High-throughput screening; human health hazard prioritization values; H3PV; risk assessment; risk screening

1. INTRODUCTION

Today it is estimated that there may be over 80,000 chemicals in commerce and the environment.⁽¹⁾ Most of these chemicals lack authoritative reference values or risk screening values, meaning their hazards are not well understood, and they may not be regulated. Although having authoritative

reference values for as many chemicals as possible would be ideal, risk assessment is a time-consuming activity with need of data, methods, and models sufficient to estimate hazard and dose response. With so many chemicals lacking assessments and authoritative reference values, the question turns to which chemicals should an agency assess first? Risk screening and human health hazard prioritization values provide regulatory agencies a means to prioritize chemicals for further assessment based on potential human exposure and human health hazards.

When combined with a margin of exposure (MOE) approach, risk screening values may be used for prioritizing chemicals based on real-world exposure. Under the MOE approach, a risk screening

¹U.S. Army Engineer Research and Development Center, Research Triangle Park, NC, USA.

²Oak Ridge Institute for Science and Education assigned to the U.S. Environmental Protection Agency, National Center for Environmental Assessment, Research Triangle Park, NC, USA.

*Address correspondence to Lyle D. Burgoon, Ph.D., U.S. Army Engineer Research and Development Center, Research Triangle Park, NC 27711, USA; lyle.d.burgoon@usace.army.mil.

value (in units of mg/kg-day or mg/m³ for oral or inhalation, respectively) would be divided by a factor, for instance, 100 in this example, to obtain a safe margin below which adversity is not expected to occur within a population. Once a chemical exposure is estimated within the MOE (e.g., within 100× of the risk screening value), that chemical would be prioritized for further risk assessment.

Computational toxicology methods provide a means to begin developing risk screening values. *In vitro* quantitative high-throughput screening assays (qHTS) provide concentration–response data. When coupled with adverse outcome pathways (AOPs; note that AOPs are not chemical-specific), these data may be able to infer adverse outcomes. By using concentration–response analysis, we can identify surrogate internal concentration points of departure, or we may be able to use the concentration–response curve to calculate risk-specific concentrations (e.g., the dose at which 1:1,000 or 1:10,000 risk of adversity occurs). Using reverse dosimetry we can convert from these surrogate internal concentrations to estimates of the external dose/concentration in units such as mg/kg-day or mg/m³ (note that external dose/concentration are environmental exposure levels).

One of the primary challenges associated with using qHTS data to inform risk assessment is the lack of reverse dosimetry data and models. Until this can be rectified, we see little recourse except to use a relative ranking scheme based on *in vitro* surrogate internal concentration points of departure or risk-specific concentrations (we term these human health hazard prioritization values or H³PV). The downside is that these surrogate values cannot be used in an MOE analysis. However, it still allows for the chemicals to be ranked relative to one another for prioritization purposes. When combined with real-world exposure-based ranking of chemicals (e.g., site-specific ranking of chemicals based on concentrations in various media), we can combine this information for ranking and prioritization purposes.

We are focused on developing methods that focus on translating nascent data streams (e.g., omics, qHTS, data mining of the vast biomedical database universe, computational toxicology) into H³PV, risk screening, risk assessments, and, ultimately, risk management decisions. This will require efforts for hazard identification, concentration–response analysis, and reverse dosimetry.

In this article, we outline a case study focused on methods for using *in vitro* qHTS data for hazard identification and concentration–response analysis. Our case study is focused on benzo [k]fluoranthene(B[k]F), a polycyclic aromatic hydrocarbon.

This work was inspired in large part by other efforts to perform high-throughput risk assessments.⁽²⁾ We concur with Judson *et al.* (2011) that pathways are key in anchoring the qHTS assays, and that reverse dosimetry is the key to moving from *in vitro* concentrations to external dose/concentration. However, we differ in several respects that will be explored further in this work. As illustrated here, we believe that risk assessors need only be concerned with the minimal set of key events (and their associated assays) that are sufficient to infer the adverse outcome. In addition, our work is focused on moving towards a semi-automated approach, where computers are used to integrate and perform a first-pass analysis of the data. Risk assessors then will utilize these analyses to facilitate their own analyses, help tune the algorithms to improve their performance, and ultimately help risk assessors increase their assessment efficiency and throughput. Furthermore, although we agree with Judson *et al.* that utilizing qHTS data in risk assessment is the ultimate goal, we have turned our focus to the near-term goal of using *in vitro* qHTS data for risk screening and hazard-based chemical prioritization.

We utilize an AOP ontology we developed to facilitate the use of first-order logic to identify hazards associated with B[k]F based on putative AOPs. We then use the ontology to help us identify which assays associated with AOP key events that are sufficient to infer adversity to use for concentration–response analysis, point of departure, and risk-specific *in vitro* concentration estimation. Future case studies may focus on reverse dosimetry to inform the generation of a risk screening value.

2. MATERIALS AND METHODS

2.1. Adverse Outcome Pathway Construction

The focus of this study was on hepatic steatosis and oxidative DNA damage. The putative steatosis AOP (Fig. 1) and the putative DNA-damage-induced cancer AOP (Fig. 2) were constructed based on existing disease pathway data available from

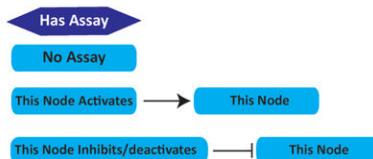
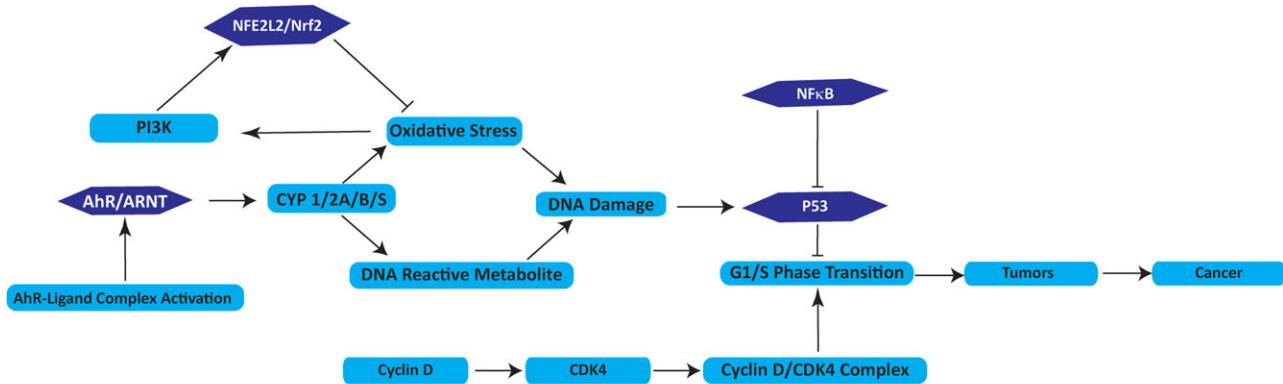


Fig. 1. Putative AOP for DNA damage leading to cancer. This AOP is a relatively well-known pathway. Each node is a key event, connected by either an arrow (representing activation) or a dash (representing repression). The hexagonal nodes are key events where we have an assay in the PubChem database. Based on the pathway and information from the literature, activation of p53 is sufficient to infer there is DNA damage. Although this is not sufficient information to infer anything about tumorigenesis or carcinogenesis, *per se*, this is important information when building an argument about the chemical's mode of action.

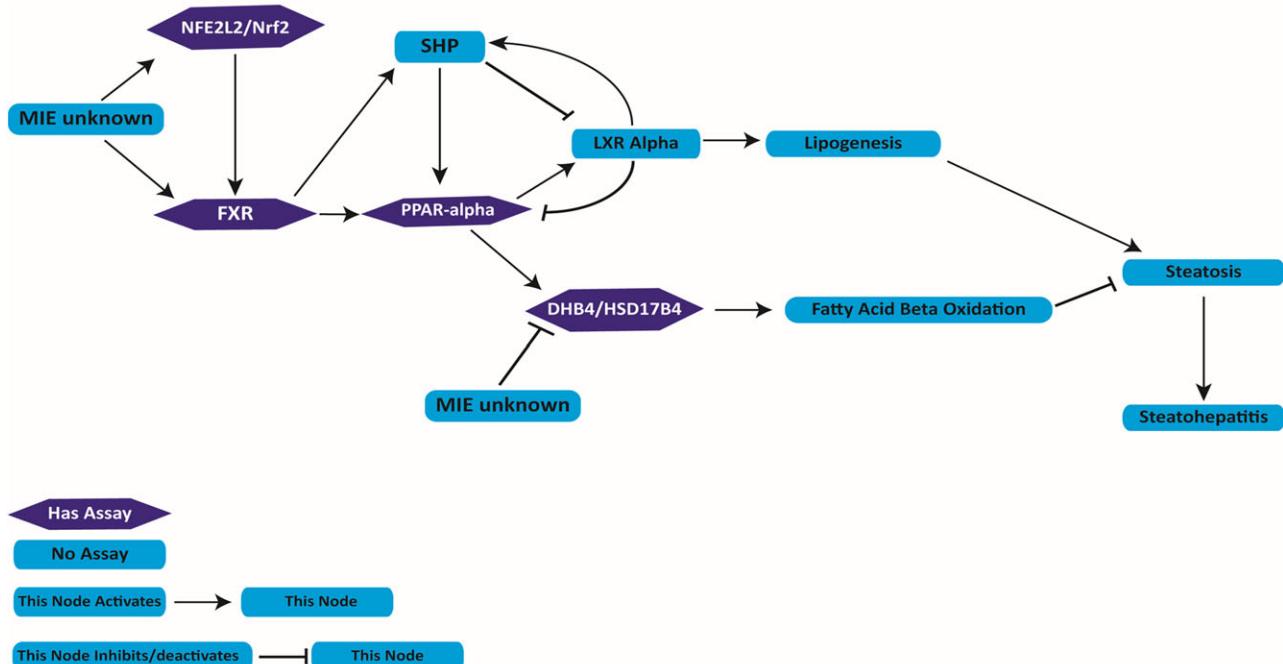


Fig. 2. Putative AOP for steatosis via decreased fatty acid beta oxidation. This AOP demonstrates two pathways that lead to steatosis. The HSD17B4 key event node is sufficient to infer steatosis based on its importance in fatty acid beta oxidation. In this case, deactivation of HSD17B4 activity will lead to decreased fatty acid beta oxidation, which will lead to steatosis. Thus, decreased HSD17B4 activity is sufficient to infer steatosis.

sources listed in the National Library of Medicine's BioSystems Database.⁽³⁾ BioSystems includes pathway data from a multitude of sources, including Reactome, the Kyoto Encyclopedia of Genes and Genomes (KEGG), and others. This information was augmented with data from literature sources to build the putative AOPs.

2.2. In Vitro qHTS Data Acquisition

In vitro high-throughput screening data were acquired from the National Library of Medicine's PubChem Database (<https://pubchem.ncbi.nlm.nih.gov/>) for all active, confirmatory assays for B[k]F under the chemical ID: 9158. We narrowed the list of assays down to just those that represented key events in our putative AOPs. For concentration-response analysis, these were narrowed to those assays that were linked to key events that were sufficient to infer an adverse outcome.

All key events within an AOP are, by definition, necessary. However, not all key events are sufficient to infer the adverse outcome within an AOP. Rothman⁽⁴⁾ defined sufficient as “[a] cause which inevitably produces the effect.” Under Rothman's sufficient-component cause model, there is at least one, and possibly more components that jointly, are sufficient to infer a disease or adverse outcome. This means that removal of just one component is sufficient to prevent the disease or adverse outcome.^(4,5) This can also be termed minimal sufficiency. By definition, if a downstream key event in an AOP is affected by a chemical (e.g., activated or deactivated), then the upstream events had to occur. Thus, if we have identified a sufficient key event and activation of this key event is sufficient to infer adversity, then we can infer that the adverse event will occur.

2.3. Concentration-Response Analysis

For the p53 putative AOP key event, we had five replicates across largely similar doses under the PubChem BioAssay IDs: 651631 (three replicates), 651743 (two different PubChem substance IDs). We integrated the evidence/data (e.g., concentration-response data) across these five replicates using a bootstrap natural spline-based meta-regression. We performed 5,000 iterations and calculated the bootstrap median-concentration response curve and the 95% confidence region. Note that we could have just as easily used a bootstrap loess meta-regression method rather than the bootstrap natural spline-

based meta-regression. Our reason for choosing the natural spline approach is one of convenience, whereby the model will be linear at the edges (this is useful if we had to perform low-dose/concentration extrapolation). However, for the sake of interpolation, we estimate the results would be similar between loess and the natural splines.

For the HSD17B4 putative AOP key event, we had one replicate (AID: 893). We observed that at the higher doses the response percentage dropped below -100%, and at the highest dose dropped below -400%. This assay is an *in vitro* enzymatic assay, meaning it is cell-free. The reference point in this assay is an untreated control, where HSD17B4 enzyme is placed in a well with the electron donor beta-hydroxybutyryl coenzyme A and the electron acceptor NAD+. Thus, we find it difficult to believe a response less than -100%.

Consider the following: the control fluorescence when the HSD17B4 is placed in the well with the beta-hydroxybutyryl coenzyme A and NAD+ is 1,000 units. A +50% increase in activity would be equivalent to 1,500 units. Likewise, a -50% activity response would be a decrease by 50%, for a total of 500 units. If after adding a chemical, the fluorescence was 0 units that would be a -100% response rate (or 100% decrease in activity). Thus, to have a less than -100% response rate, the fluorescence assay would have to be reading negative values, which is likely if the fluorescence values are extremely low, and likely occurring after various data transformations are applied. Thus, as these values are biologically and analytically impossible/improbable, we have censored the data whose values are <-100% (i.e., -400%) to be -100%.

2.4. Point of Departure Estimation

We estimated the point of departure as the concentration at which the upper 95% confidence interval demonstrated deviation from the “background” rate near the low-concentration asymptote. We analyzed the slope to identify the point at which the curve began to depart from the asymptote.

2.5. Risk-Specific *In Vitro* Concentration Estimation

We assumed that the maximal asymptote represents maximal risk, and that the minimal asymptote represents the baseline risk. For instance, if a risk assessor were interested in identifying the dose that

causes a risk of 1:1,000 (where maximal response is considered 100% risk), we would identify the maximum response and divide that by 1,000, to yield the response factor. We then analyzed the region of the curve less than the point of departure to identify the maximum value for the upper 95% confidence interval using a small interval (i.e., we would focus in on a narrow concentration range less than the point of departure with a small interval between the concentrations that were input to the model). We identified the risk-specific *in vitro* concentration as the concentration from the model that yielded the value closest and less than the response factor (in other words, if the response factor is 10%, then we would identify a concentration that yields a response as close to and less than 10% as possible). If the sum were between two concentrations, we used the lower concentration in order to be public health protective.

2.6. Analysis Environment

All statistical analyses and modeling were performed in R (version 3.1.1). The ontology is encoded using the Web Ontology Language (OWL). We chose to utilize OWL as it supports necessary, sufficient, and necessary and sufficient conditions. Computational reasoners can be used to draw these same conclusions using the ontology, and to automate the process. Because of the simplicity of the AOPs and the sufficiency relationships, we did not need to use a computational reasoner to make inferences.

3. RESULTS

3.1. Adverse Outcome Pathways

We constructed two putative AOPs based largely on existing disease knowledge. The p53-tumorigenesis and cancer putative AOP is based on existing pathway information, including WikiPathway RB in Cancer (<http://www.wikipathways.org/index.php?title=Pathway:WP2446&oldid=76353>), KEGG Cell Cycle (http://www.genome.jp/kegg-bin/show_pathway?hsa04110), and the Reactome Cell Cycle Checkpoints (<http://www.reactome.org/PathwayBrowser/#DIAGRAM=69620&PATH=1640170>) pathways (Fig. 1). Fig. 2 depicts the HSD17B4-steatosis putative AOP, based on the Reactome Peroxisomal Lipid Metabolism Pathway (<http://www.reactome.org/PathwayBrowser/#DIAGRAM=390918>), and other articles.^(6,7) Note that the pathway data sources are curated based on the literature.

3.2. Identification of Sufficient Key Event–Adverse Outcome Relationships

We identified key events upstream of the adverse outcome that are sufficient to infer the adverse outcome. If we consider the p53-tumorigenesis putative AOP (Fig. 1), activation of p53 is sufficient to infer DNA damage. Tumor protein p53 is a protein critical in cell-cycle regulation and thus functions as a tumor suppressor. In a normal healthy cell, inactive p53 is tagged for ubiquitylation by the E3 ubiquitin ligase, MDM2, and degraded by the 26S proteasome,⁽¹⁾ making baseline protein levels of the p53 protein low. When there is DNA damage, p53 is phosphorylated, preventing ubiquitination, and activated, leading to cell-cycle arrest and allowing the damaged DNA to be repaired.⁽⁸⁾

Although having a nonfunctional p53 may be necessary for tumorigenesis in some cases, it is not sufficient to infer that tumorigenesis will occur, as that would also require cell-cycle activation. Thus, for our purposes, we will focus on the inferential relationship where p53 activation infers DNA damage.

In consideration of the HSD17B4-steatosis putative AOP, we know HSD17B4 is a key protein in fatty acid beta oxidation and that a lack of HSD17B4 activity is sufficient to infer steatosis.⁽⁷⁾

3.3. Hazard Identification Using *In Vitro* qHTS Data

We obtained assay data from PubChem for the hexagonal AOP key events (Figs. 1 and 2). In the p53 putative AOP, these include assays for AhR/ARNT, Nrf2, NF- κ B, and p53. For the HSD17B4-steatosis putative AOP, we have assay data for Nrf2, FXR, PPAR-alpha, and HSD17B4. However, based on the sufficiency arguments we have encoded into the ontology, we will only focus on the p53 and HSD17B4 assays. Thus, based on the sufficiency relationships, we infer that there is DNA damage present if p53 is activated, and we infer that steatosis occurs when HSD17B4 activity is decreased.

For the p53 activation key event, we identified five replicates from the PubChem database (three are from the same data set). Based on the results of our bootstrap natural spline-based meta-regression (Fig. 3), we have identified that the replicates are in good agreement, demonstrating a concentration-dependent activation of p53. The data are approaching a plateau near 100 μ M. Based on the assay data,

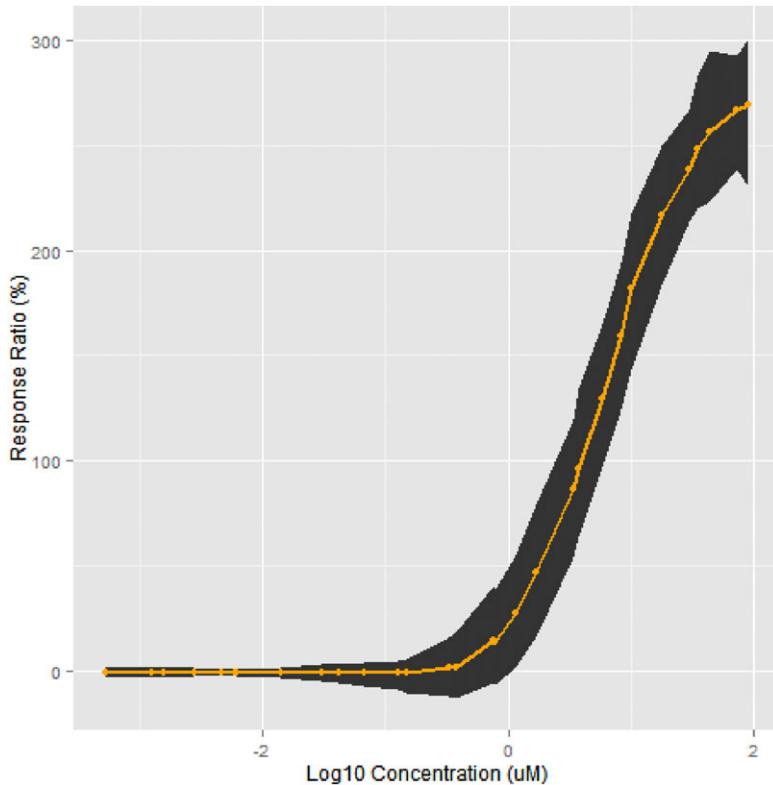


Fig. 3. Bootstrap natural spline-based meta-regression of the concentration–response curve for the p53 qHTS assay data. The concentration–response data from five assays were integrated using bootstrapping (5,000 iterations) and natural splines to perform meta-regression. The line represents the median concentration response across the 5,000 iterations. The shaded region surrounding the median line represents the 95% confidence interval.

it appears that there is a threshold below which p53 is not likely to be activated. Thus, there is evidence that DNA damage is occurring following B[k]F exposure; however, there appears to be a threshold below which DNA damage is not inferred, based on lack of p53 assay activation. We have moderately high confidence in this conclusion, based on concordance between multiple replicates of the p53 assay for this chemical.

We identified only one replicate from the PubChem database for HSD17B4 activation/inhibition. The concentration–response curve is depicted in Fig. 4. The data appear to show a threshold below which HSD17B4 activity remains relatively steady. At higher doses of B[k]F, the HSD17B4 activity rapidly declines. This suggests that at higher concentrations of B[k]F exposure, HSD17B4 activity has decreased, and that steatosis is likely. Thus, we have evidence of steatosis following B[k]F exposure, with a threshold below which steatosis is unlikely. We have low confidence in this conclusion, and low confidence with regards to the threshold at which effects are unlikely to occur, because only a single replicate of the HSD17B4 assay was available.

3.4. Point of Departure and Risk-Specific Concentration Estimation

We define the point of departure as that concentration where one can reasonably anticipate deviation from the control response, as measured by a marked deviation of the curve away from the asymptote. This is the point at which the slope of the curve increases (decreases) markedly and denotes the curve’s departure from the asymptote in a monotonically increasing (decreasing) curve. For p53 we estimate an internal concentration point of departure of $0.751 \mu\text{M}$ based on the deviation of the slope from baseline. This method for point of departure estimation is based on the fact that at low doses the concentration–response curve approaches an asymptote. We define baseline responses as those that occur where the slope has minimal changes, typically associated with measurement/assay noise. In other words, this region that approaches the asymptote is not expected to be discernable from control responses.

This point can be easily seen in a plot of the slope (Fig. 5), where $0.751 \mu\text{M}$ exhibits the first largest slope increase (44.4% response ratio units/ μM).

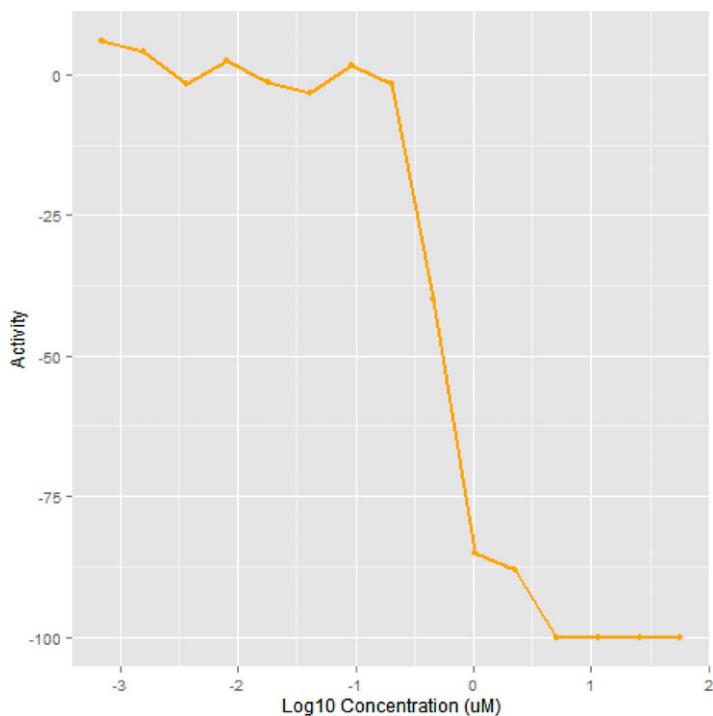


Fig. 4. Concentration–response curve for the HSD17B4 qHTS assay. Only a single assay was available for this analysis, so no bootstrapping or regression was performed. This plot depicts the concentration–response for the only available replicate.

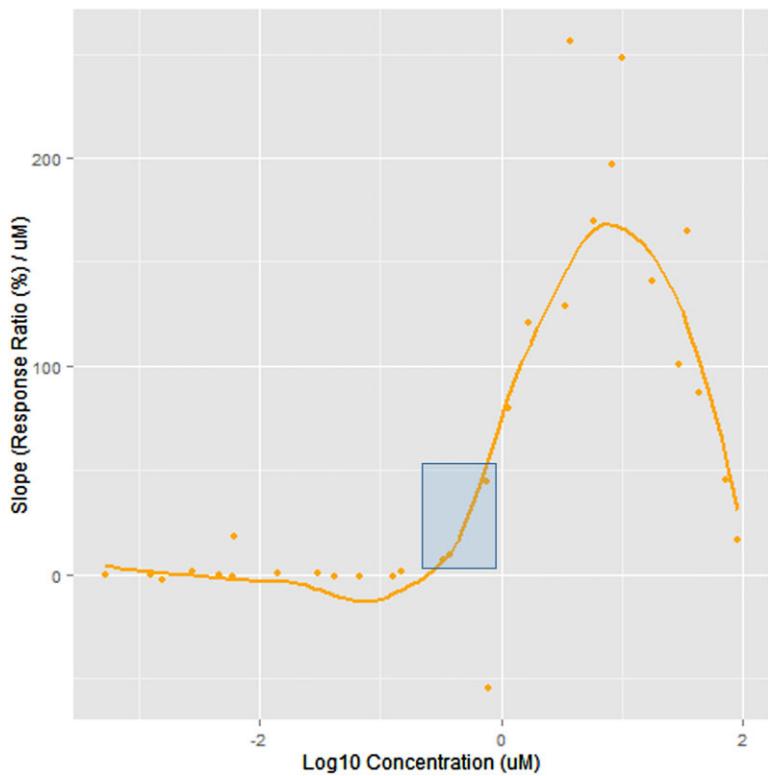


Fig. 5. Concentration–response slope graph for the p53 qHTS assay data. This plot can be helpful in determining the point of departure when used with the concentration–response plot. Based on this plot, it would appear that the concentration–response curve is beginning to deviate from the baseline from 0.337 to 0.751 μM (boxed area). The slope has definitely increased by 0.751 μM , followed by a sharp decrease in slope at 0.782 μM (due to an unusually small difference in the concentrations in the denominator). Matching this with the concentration–response plot in Fig. 3, it becomes clear that the slope increase from 0.337 to 0.378 μM is within the normal variation for the baseline; however, the slope increase from 0.378 to 0.751 μM is large, and helps support the case that 0.751 μM is a reasonable point of departure.

Because of the use of secant lines (i.e., lines intersecting two points on the curve) to calculate the slope, there is an anomalous, sharp decrease in slope at $0.782 \mu\text{M}$. This is due to the small increment in the denominator gaining greater weight over the small decrease in response ratio. However, by analyzing the curve (Fig. 3) and the slope (Fig. 5) simultaneously, it is clear that this is a small anomalous blip in the slope curve that holds little weight in the overall interpretation.

We can use the modeled concentration-response curves to also calculate risk-specific *in vitro* concentrations. A risk-specific *in vitro* concentration is a concentration at which a certain percentage of response, compared to maximal response, will occur. In the p53 example, we assume that maximal risk (near 100%) is equivalent to the increasing response asymptote (higher concentrations). We also assume that background risk is equivalent to the decreasing response asymptote (i.e., the response at lower concentrations near the asymptote is assumed to be at background levels for an activator). We recognize that this is a simplifying assumption, and that the *in vitro* cells in culture likely replicate a single genotype. This assumption is important as it means that we are not able to account for population variability.

The 1:10,000 and 1:1,000 risk *in vitro* concentrations are approximately concentration $0.28 \mu\text{M}$ and $0.29 \mu\text{M}$, respectively. These concentrations are approximately 40% lower than the point of departure, within the asymptote.

For HSD17B4, we were not able to calculate a reliable point of departure or risk-specific concentration. With essentially a single replicate, we do not feel comfortable calculating these estimates.

4. DISCUSSION

Rapid risk screening methods are necessary to begin to tackle the risk assessment challenges posed by the more than 80,000 chemicals in commerce and the environment. Risk screening will help prioritize chemicals for future research and assessment efforts. By focusing resources in this way, we can begin to take more reasonable steps to address this challenge.

In addition, risk screening assessments can provide critical toxicological information for emergency response when a more traditional risk assessment is lacking. Even in an emergency situation, a data limited case, such as our B[k]F steatosis case study, can be helpful. Although in the steatosis case study

we only had one replicate concentration-response curve, this information is far more helpful in establishing a potentially safe level as compared to having no data at all. For instance, consider a case like the 4-methylcyclohexanemethanol (MCHM) spill in West Virginia in January 2014. There, environmental public health agencies lacked an authoritative toxicological assessment of MCHM. Eventually a subacute study from the manufacturer was identified. However, what if no values were found that could be used? Risk managers would welcome a single replicate qHTS assay, such as our steatosis case for B[k]F, over no data whatsoever.

Rapid risk screening will require rapid toxicity and hazard assessments, which includes rapid hazard identification and rapid dose-response assessments. In this article, we describe a case study that lays out our methods to perform the rapid hazard identification and rapid concentration-response analysis. Our approaches are readily amenable to automation. For instance, we can automate the hazard identification process by (1) leveraging an ontology that maps assays to AOP key events with hand-curated sufficiency relationships that allow us to infer adverse outcomes from one or more sufficient key events, (2) statistical methods that can determine if an assay is active at various concentrations, and (3) combining the statistical information with the ontology and utilizing first-order logic to identify those concentrations where an adverse outcome is likely. The concentration-response analysis is a semi-automated approach, where the bootstrap spline-based meta-regression can be easily fully automated, and we anticipate humans would make the final decision on points of departure and risk-specific concentrations.

In applying these approaches, we have begun to discuss our confidence in the results. Creating a confidence framework is still an ongoing process for us. As risk assessors, we feel it is important to articulate our confidence in our conclusions using these data. Given our experience with data-poor chemicals, and real-world situations where risk management decisions need to be made in the absence of good-quality data, we feel it is important to be able to use all data that are available to us. Thus, to us, confidence is a continuum from low to high. We have the lowest confidence in data from a single replicate. We have the highest confidence in data from multiple orthogonal assays where each has multiple replicates. We have significantly fewer reservations broadcasting low-confidence hazard identification calls than low-confidence point of departure calculations, reserving

low-confidence point of departure calculations for those cases where it is absolutely necessary.

In our p53 case study, we were able to integrate evidence from five assays. Based on our semantic/ontology analysis, we were able to identify that B[k]F likely causes DNA damage. This is based on the semantic relationship between p53 and DNA damage, such that p53 activation is sufficient to infer DNA damage has occurred. We estimated a POD for DNA damage, as measured by the p53 assays, of $0.751\mu\text{M}$, with a 1:1,000 and 1:10,000 risk at $0.29\mu\text{M}$ and $0.28\mu\text{M}$, respectively.

Our confidence in the results would have been increased had orthogonal assays (i.e., assays measuring the same key event in different ways) been available, and if different groups had performed the assays. However, given the reproducibility across the five assays, across different spans of time, we feel moderately confident in these results.

There are limitations and assumptions associated with basing the 1:1,000 and 1:10,000 risk levels on maximal response. For instance, this assumes that maximal risk is equated to maximal response, which may or may not be true. It also assumes that all response rates in any given assay have some inherent risk, which may not be true in all cases. Furthermore, for risk to be equated to response, we would have to presume that cellular and molecular responses in these assay systems are binary in nature—where graded responses are the result of some percentage of cells being active (e.g., 90% response is equivalent to 90% of the cells in a system being active). This is why it is important that decisionmakers and investigators understand the strengths and limitations of the assay systems, and that assessments be developed by people who understand the analysis approaches and assays. The 1:1,000 and 1:10,000 risk analyses are demonstrated here as a proof of concept for a more quantitative risk approach than the POD method; however, the results should be taken with care when using them in a risk screening environment.

We also assessed the ability of B[k]F to cause steatosis through decreased HSD17B4 activity. Based on our semantic/ontology analysis, we were able to identify that B[k]F could cause steatosis, by decreasing HSD17B4 activity. We have little confidence in this result due to having only one data set available. Because we lacked additional assay data, we were not comfortable calculating a point of departure or risk-specific concentrations. This represents a case where risk assessors might request additional assays be run to increase the con-

fidence in the HSD17B4 results. As our ontology and literature analysis identified HSD17B4 as being sufficient to infer steatosis, we would only need additional data for HSD17B4, preferably with orthogonal assays. Requesting data for additional key events, which were not sufficient to infer steatosis, would have very little value for our hazard identification and concentration-response analyses.

Leveraging Rothman's idea of minimal sufficiency⁽⁴⁾ allows us to quickly hone in on those AOP key events that are the most important and informative for risk assessment. Under minimal sufficiency, we need only monitor and test those key events that are sufficient to infer an adverse outcome. Ideally, orthogonal assays would be available. This means that only a small number of key events across the AOP universe will actually require testing/monitoring, further driving down the number of required tests, while maximizing the information gain.

What we have described here is just the first step in a more comprehensive risk screening paradigm. Using these approaches, we can rank and prioritize chemicals for further research and assessment in a semi-automated fashion (Fig. 6). This semi-automated work would serve as the input to a gating function, where chemicals would be continuously prioritized/reprioritized by regulatory agency decisionmakers based on the identified hazards, PODs, or risk-specific concentrations. Prioritization could also be based on the availability of data. For instance, we currently lack the information and models required for reverse dosimetry for many chemicals. We could perform reverse dosimetry for chemicals, based on priority order, for those chemicals where the data are available. In other words, chemicals that have data that can be used for reverse dosimetry would be allowed through the gate before chemicals with higher priority that lack these data. Following reverse dosimetry, we would have PODs and risk-specific dose/concentration levels in mg/kg-day or mg/m³. This information could be used in a risk screening context to further prioritize chemicals for additional research and risk assessment.

In addition, because risk assessors will likely lack the ability to perform reverse dosimetry for most chemicals, we believe that relative ranking of chemicals based on the surrogate internal concentration may be the best alternative for chemical prioritization at this time. By combining these relative rankings with site-specific or nationwide exposure-based chemical rankings, we can calculate a relative

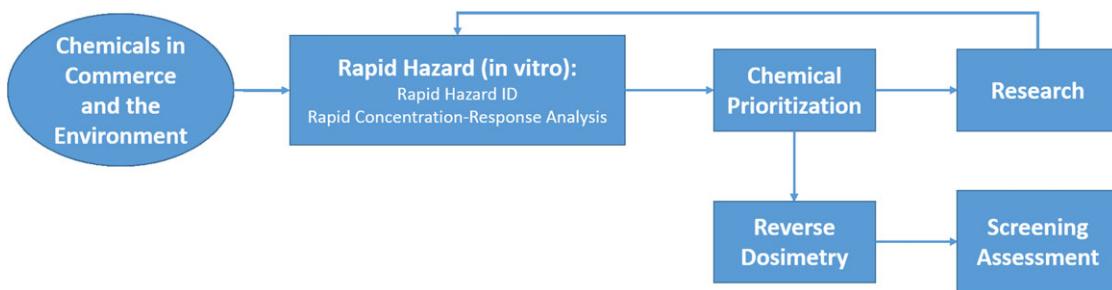


Fig. 6. Overview of rapid hazard and rapid risk screening. This graphical overview illustrates the information flow starting at chemicals in the environment, and ending at a screening assessment. This overview is specific for *in vitro* qHTS data. We envision using the semi-automated methods described in our case study in the rapid hazard (*in vitro*) step. By making this process semi-automated, we anticipate being able to prioritize hundreds to thousands of chemicals in the relatively near future (less than three years). The principal bottleneck at this stage is the availability of putative AOPs, “ontologizing” them, and identifying sufficiency relationships. As chemicals are prioritized, we can identify those that have the most value in having additional research performed. As data and models become available to perform reverse dosimetry, we can move chemicals out of the gate, perform reverse dosimetry, and develop screening assessments.

risk prioritization based on either the rank order of the sums or products of the ranks. In other words, the overall risk prioritization would be the product of the chemical's relative ranking based on surrogate internal concentration and the chemical's relative ranking based on site-specific or nationwide exposure. Although this lacks the elegance of an MOE approach, it still has utility in prioritizing chemicals for further assessment. Our case study demonstrates that quantitative evidence integration can be accomplished using fairly routine methods (e.g., bootstrap natural spline-based meta-regression), but that we require criteria for determining our level of confidence in the data in a transparent way. In addition, criteria for determining causality of an adverse outcome will be helpful to ensure we are making clear, logical, and transparent arguments. For instance, we have little doubt others would disagree that there is little confidence in stating that B[k]F causes steatosis based on the available qHTS data. However, without clear and transparent criteria one could easily make an argument that there is strong/moderate/weak evidence concerning DNA damage based on the p53 assays at varying doses without providing a clear, convincing, and logical case. Use of a confidence framework would allow all involved (e.g., stakeholders, risk assessors, policymakers) to understand the criteria that were applied, and how they were applied.

5. CONCLUSIONS

We have developed a case study that lays out a framework to move qHTS data into risk screening. In the near future we will migrate our case-

study-specific code into semi-automated methods to facilitate our analyses in R. We feel it is important to keep risk assessors active at the most important steps—determining a point of departure, or identifying risk-specific concentrations. We anticipate that by adopting a semi-automated process we will be able to facilitate transparency throughout the analysis process, from data acquisition and screening through risk screening assessment.

We learned several lessons in the development of this case study. For instance, we have identified that a confidence framework needs to be created to guide assessors in describing the level of confidence that the data confer for specific adverse outcomes. In addition, a causal framework will facilitate the analysis of causal relationships between chemical exposure and adverse outcome using these emerging and nascent data streams (e.g., qHTS, omics, computational toxicology, data mining). These are but a few of the future directions we are currently engaged in.

Overall, we feel confident that as more qHTS data become available we will be able to translate these into risk assessment research needs and risk screening assessments. We are equally confident that transparency can be enhanced over time, as we continue to migrate toward more computationally efficient, semi-automated methods.

ACKNOWLEDGMENTS

The authors thank Drs. Ila Cote, Michelle Angrish, Reeder Sams, and John Vandenberg for their comments on earlier drafts of this article. This project started when Dr. Burgoon was at the U.S. Environmental Protection Agency's National Center

for Environmental Assessment, and completed after Dr. Burgoon began working at the U.S. Army Engineer Research and Development Center. The U.S. Environmental Protection Agency (EPA) Human Health Risk Assessment Program and the U.S. Army Environmental Quality and Installations Rapid Hazard Assessment Focus Area Research Project provided financial support. Drs. Druwe and Yost and Mr. Painter were supported by an appointment to the Internship/Research Participation Program at the Office of Research and Development, U.S. EPA, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. EPA.

REFERENCES

1. Kavlock RJ, Ankley G, Blancato J, Breen M, Conolly R, Dix D, Houck K, Hubal E, Judson R, Rabinowitz J, Richard A, Setzer RW, Shah I, Villeneuve D, Weber E. Computational toxicology—A state of the science mini review. *Toxicological Sciences*, 2008; 103(1):14–27.
2. Judson RS, Kavlock RJ, Setzer RW, Cohen Hubal EA, Martin MT, Knudsen TB, Houck KA, Thomas RS, Wetmore BA, Dix DJ. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. *Chemical Research in Toxicology*, 2011; 24(4):451–462.
3. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. The NCBI BioSystems database. *Nucleic Acids Research*, 2010; 38(Database issue):D492–D496.
4. Rothman KJ. Causes. 1976. *American Journal of Epidemiology*, 1995; 141(2):90–95; discussion 89.
5. Höfler M. The Bradford Hill considerations on causality: A counterfactual perspective. *Emerging Themes in Epidemiology*, 2005; 2(1):11.
6. Kay HY, Kim WD, Hwang SJ, Choi H-S, Gilroy RK, Wan Y-JY, Kim SG. Nrf2 inhibits LX α R α -dependent hepatic lipogenesis by competing with FXR for acetylase binding. *Antioxidants & Redox Signaling*, 2011; 15(8):2135–2146.
7. Reddy JK III. Peroxisomal β -oxidation, PPAR α , and steatohepatitis. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, 2001; 281(6):G1333–G1339.
8. Gilbert S. *Developmental Biology*, 10th ed. Sunderland, MA: Sinauer Associates, Inc., 2013.