Open Peer Review on Qeios

Improving the Integration of Epidemiological Data Into Human Health Risk Assessment: What Risk Assessors Told Us They Want

Carl Phillips, Igor Burstyn, David Miller, Ali K Hamade, Raghavendhran Avanasi, Denali N. Boon, Jennifer Reed, Saumitra V Rege, Sandrine Deglin

Funding: Fully funded by the Health and Environmental Science Institute (HESI)

Potential competing interests: The authors declare no financial conflicts of interest. IB and CVP disclose activism on behalf of better bias analyses and other improvements in the conduct of epidemiology, belief in existence of thresholds in biology, and sincere desire to make risk assessment more quantitative.

Abstract

The practical contribution of epidemiology studies of chemical exposures and disease outcomes is to inform risk assessment decisions. But risk assessors are widely dissatisfied with the conduct of epidemiology in this area. There is a widespread feeling the epidemiology is neither done nor reported in a way that is informative to them. To help identify the points of dissatisfaction and possible areas for mutual learning, we conducted a survey of risk assessors, seeking their opinions of epidemiology research. We present a few quantitative measures and a thematic analysis of responses to open-ended questions. Many risk assessors (with some adamant exceptions) believe that epidemiology has great potential to contribute to risk assessment, but falls short in many ways. Respondents identified inexcusable omissions (e.g., failing to clearly define and report the exposure measurement) and straightforward design choices (e.g., failure to stratify by exposure level) that render epidemiology result basically useless to them. Respondents also brought up a wide collection of more complicated and subtle concerns that could lead to further improvement of useful results. We identify areas where mutually-educating interdisciplinary dialog seems particularly promising. Epidemiology research is expensive, and risk management decisions more so, therefore it is inexcusable to not try to make epidemiologic research more useful for informing decisions.

Carl V Phillips[1], **Igor Burstyn**[2,#], **David Miller**[3], **Ali K Hamade**[4], **Raghavendhran Avanasi**[5], **Denali Boon**[6], **Jennifer Reed**[7], **Saumitra V Rege**[8], and **Sandrine Deglin**[9]

[1] *Epiphi Consulting*

[2] *Environmental and Occupational Health, Drexel University*

[3] *U.S. Environmental Protection Agency (retired)*

[4] *Oregon Health Authority*

[5] *Syngenta Crop Protection*

[6] *Gradient Corp.,*

[7] *Bayer Crop Science*

[8] *ExxonMobil Biomedical Sciences Inc.*

[9] *Health and Environmental Sciences Institute*

[#] Corresponding author: igor.burstyn@drexel.edu

Introduction

To what extent is epidemiology, as currently practiced, perceived as useful for human health risk assessment?

Scientific research has many purposes, ranging from discovering properties of nature to satisfy curiosity and imagination, to providing dry quantitative details that are needed to optimize a policy decision. Because of the ephemeral nature of what it measures (for a particular population and time that will never again exist), epidemiology is almost all solidly at the latter end, with only a bit of work in the middle ground that includes such inquiries as “does this exposure apparently cause a particular outcome at all, ever?” Once that question has been answered, further “hazard identification” analyses that do not provide useful quantitative estimates are not informative. Thus, epidemiology studies of particular chemical exposure and disease combinations really have just one job: to estimate how much harm particular exposure levels are causing, allowing society to make informed decisions about controlling the exposure.

If epidemiology in this area is not done in a way that can quantitatively inform decision-making about risk then it is, at best, not useful. Grumblings to that effect can frequently be heard from risk assessors. Presumably epidemiologists would like their work to be useful and used, and risk assessors should definitely prefer this. But there is remarkably little interchange between these groups to try to make it happen (Deglin et al., 2021). Epidemiologists in academia or government research institutes usually design and report the results of studies using whatever methods and tools they happen to be familiar with, and often fail to know or seek to know what would be most useful. Risk assessors then fatalistically lament that the epidemiology is usually not very useful.

It does not have to be this way.

To offer the start of a path toward remedying this wasteful disconnect, the Health and Environmental Science Institute (HESI) conducted a survey of risk assessors, about their perceptions of epidemiology in their area, how useful it is in meeting their needs and wants, and how it could be made more useful.

Previous literature

Most scholarly writings on this topic are a collection of aspirational statements about what epidemiology can contribute, either in a narrative format (Samet et al., 1998) or checklists of suggestions, considerations, or “criteria”, often for ex post evaluation of studies rather than to guide study conduct (e.g. The London Principles (Federal Focus Inc., 1995); the Matrix (Burns et al., 2019; LaKind et al., 2020; LaKind et al., 2023); (Goodman et al., 2020)). Some of the previous efforts communicated the opinions and preferences of people working in risk assessment, but only a handful (the Matrix used a

workshop of four people, Samet et al. had one co-author who worked in risk assessment for government). The previous literature thus amounts to abstracts of the lore contained in textbooks and what a few experts think should be done, and lacks the important step of listening to practicing risk assessors who have seen the epidemiology and would like to tell epidemiologists “what you want, what you really, really want” (The Spice Girls, 1996).

Samet et al. is mostly an introduction to risk assessment, presenting one version of the basics of the concept and mechanics. The paper includes a few pages of idealized notions of what epidemiologists should try to do, but it is not clear if readers would learn what they needed to better operationalize. LaKind and Burns, with various coauthors, presented the series of Matrix papers addressing these questions with an emphasis on epidemiologic practice and guidelines. The Matrix was designed as a focal point of conversation between risk assessors and epidemiologists to encourage greater communication about mutual needs and wants, but even its creators have expressed concern (with good reason) that those attending to it will treat it as a mere checklist.

It is difficult to imagine that anyone ever got better at doing valid and useful epidemiology via checklists, in this area or any other. These lists are invariably an unsorted mix of some bits of good advice that are too abbreviated to be useful unless the audience already understands it, demands from conformity that are either obvious or unnecessary, and conflation of must-do prescriptions and vague aspirational goals. They usually also include a scattering of flat-out harmful recommendations and claims (e.g., a study that is all about measurement should have a pre-specified hypothesis; a population representative sample is always best). Perhaps even worse than the latter, they just do not provide any advice about what really needs to be done differently. It is easy to imagine a conversation following from one of these along the lines of:

> *Epidemiologist: “Yay me! I checked those ten boxes.”*
> *Risk Assessor: “Eight of those were never a serious issue, and for the two that are the major issues, you checked the boxes but did not actually give us what we really really want.”*

Aspirational summary catalogs of how someone thinks a particular complex task should be done may offer newcomers some basic literacy about a topic, but they give little or no sense of what really really needs to be done better. Even a textbook-length guide would not be sufficient to bridge the gap between current practice and what risk assessors want and find in epidemiologic literature. What could help instead is transfer of experiential knowledge that comes from practice -- learning about how to do one or a few things genuinely better and then putting that into practice -- as it happens in all areas of scholarship and engineering. For this to occur, it seems necessary for the parties concerned to engage with one another in cooperative co-mentoring, avoiding an authoritative mindset. Therefore, we reverse the top-down perspective that dominates the literature and offer a client-centered approach that seeks input from the front lines of risk assessment.

Methods

We sent an online survey to a convenience sample of our target population, a combination of regulators and those who

support their work (at the US federal and state levels, and national level for various countries) along with industry employees, academics, and consultants working in risk assessment. We collected invited responses starting in September 2022, and the survey invitation then opened and we sent it out via social media starting on 27 February 2023 (a few of the responses during the first phase were also apparently snowballed via one of the invited respondents sharing the link). The survey closed on 15 April 2023.

The survey asked about the respondent's professional background and their experience using epidemiology results in risk assessment. A visual analogue scale was used to assess their perception of the potential and actual usefulness of epidemiology (see survey instrument for details). Three open-ended questions asked about the pros and cons of epidemiology practice for risk assessment, and one open-ended question asking for an example of them having used epidemiology in risk assessment.

The questions are further detailed in the results reporting, and the survey instrument is available at hesiglobal.org/hesi-environmental-epidemiology-committee-risk-assessors-survey/. The full de-identified dataset is available upon request.

The primary goal was to collect a broad range of assessments with an emphasis on open-ended responses. We did not try to ensure the population was optimal for particular quantitative measures, though we did a few tallies and comparisons when they seemed particularly robust and valid. We sought only spontaneous responses to the open-ended questions (presumably the most strongly-held views), and thus can only speculate about how many others would have endorsed a particular observation if prompted. Since the sampling properties are largely unknown, we did not attempt any analysis of the results that would depend on assumptions about sampling. We expect that our respondents are not representative of all risk assessors, tending toward the more networked and "high powered" end of that population (e.g., they are national regulators, not local zoning officials, or scientists at major multinational companies, not compliance officers from small companies), and have thought more (negatively or positively) about epidemiology, which is useful for our study goals. We strove to capture the breadth of strong views as opposed to characterizing typical or average mindset.

Results And Discussion

Study population

We received 75 completed responses (one set of open-ended responses was a group effort from a Canadian agency). A small number of answers were missing for some questions; missing and apparently erroneous responses were dropped for reporting the relevant statistics from those answers (detailed below).

About half of the respondents were government regulators or related government officials (Table 1).

Table 1. Employment and employment location of respondents

Government Regulators (includes all government agency positions and support labs/consultancies)	U.S. state government	6	
	U.S. federal government	2	
		Canada	5 (incl. one group response)
		Europe	8
		Other	1
		Employer not stated	13
Other (industry, academia, other consulting)	U.S. / Canada	22	
	Europe	7	
		Other	1
		Employer not stated	7

The educational background of the respondents, a proxy for the focus of their research and their disciplinary preferences and biases, was primarily toxicology, along with a substantial number from other natural science or engineering fields. Those who studied epidemiology were invited based on their work carrying out risk assessments (Table 2).

Table 2. Educational Background of Respondents (highest degree attained)

Toxicology	PhD	30
	Master's	8
Epidemiology	PhD	5
	Master's	3
Other (a variety of fields, almost all natural science or engineering)	PhD	18
	Master's	8

Rated potential and actual value of epidemiology in risk assessment

We included a pair of quasi-quantitative questions, asking respondents to rate what they considered the potential value of epidemiology for use in risk assessment and the actual (current) value of epidemiology as practiced, on a scale of 0 to 100. Given the inevitable lack of calibration for a rating like this, it is impossible to determine exactly what each number means and how much of the variation is due to different notions of what, e.g., "75" means. But some observations about the patterns of responses seem robust.

The responses (summarized in Table 3 and depicted in Figure 1), covered most of the possible range, in terms of both absolute value and the magnitude of the disparities between the potential and actual practice. Table 4 shows that a large portion of respondents were disappointed with the conduct of epidemiology compared to its potential, for various measures of "disappointed".

Table 3. Ratings (scale of 0 to 100) of the Value of Epidemiology in Risk Assessment (n=70)

	mean	median	interquartile range
Potential value	73	80	(60, 94)
Actual value as practiced	43	43	(20, 59)

(5 responses scored the actual as higher than the potential, which might be a transposition error, but is perhaps indicative of noise in all the data beyond our ability to assess. We omitted those entries from these results.)

Table 4. Respondents suggesting a large difference for Actual vs Potential value of Epidemiology

Comparison	Disappointed by the Current State of Epidemiology compared to its potential (out of n=70)
Actual ≧30 points lower than potential	30 (43%)
Actual ≧20 points lower than potential	52 (74%)
Actual ≧50% lower than potential	26 (37%)
Actual ≧33% lower than potential	47 (67%)

To organize the distribution of responses, we imposed a "type" categorization for the combination of responses as denoted by the regions in Figure 1:

- "Optimists" do not think epidemiology is terribly useful currently, but are optimistic that it has the potential to be substantially better (n=12, 17%).
- Pessimists do not think epidemiology is terribly useful, and it does not have the potential to be much more useful (n=14, 20%).
- "Realists" -- so labeled because they reflect the rough aggregate opinions of the present authors -- think epidemiology is already moderately useful for risk assessment, but mostly see substantial room for improvement (n=29, 41%).
- "Contented" think epidemiology is very useful as is, and thus there is little room for improvement, though most of them see some room for even greater contributions (n=15, 21%).

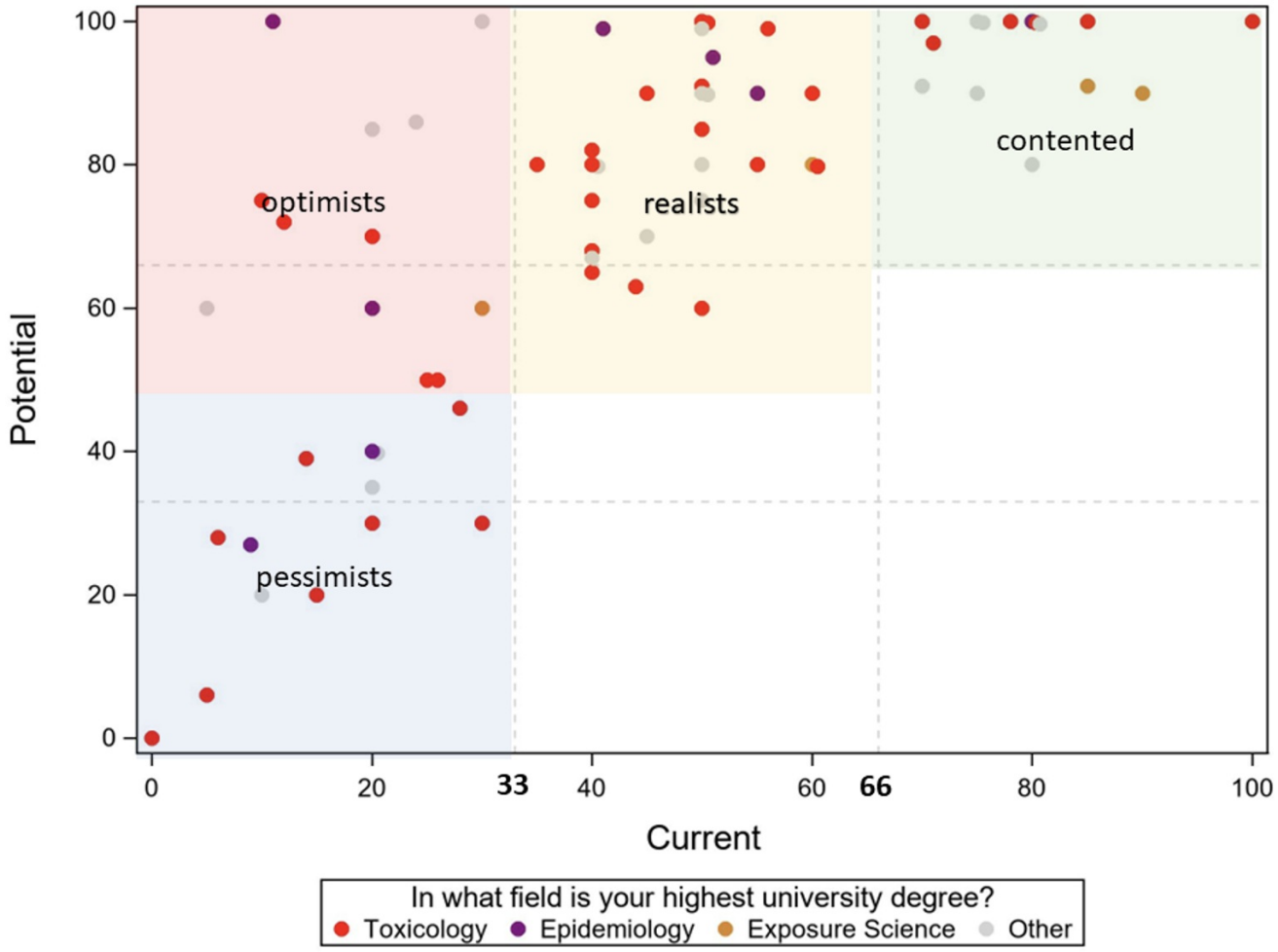

Figure 1. Perceived current and future value of epidemiology in risk assessment

The distribution of opinions in Figure 1 varies wildly, with most of the possible range represented by some respondents. It would be interesting to learn if risk assessors with different assessments would shift toward greater consensus if they sat down and discussed their views. The lack of consensus is unlikely to be an artifact of widely differing levels of sophistication, given the sample skews heavily toward the most sophisticated risk assessors. Thus, it is interesting to ask if we can identify correlates of the distribution in our data. Figure 1 illustrates the distribution by disciplinary background (highest university degree), which shows no particular patterns compared to the overall distribution, though the small number of epidemiologists tended toward dissatisfaction with current practice. There was some difference by current professional sector (Table 5), but nothing that seems substantial or meaningful given the limitations of our sample. Industry risk assessors appeared to be a bit less happy with current practice compared to government regulators, but this difference is less than we expected to see.

Table 5. Perception of the value of epidemiology “type” by professional sector

Sector	percentage of “types” across each sector			
	Pessimists	Optimists	Realists	Contented
Government (n=33)	18	15	39	27
Industry (n=20)	20	30	30	20
Academia (n=8)	38	0	50	13
Other (n=9)	11	11	67	11

We expected there would be association with responses to the three open ended questions about how epidemiology could be made more useful (discussed below), as measured by the relative quantity of positive and negative observations, vehemence, or level of detail in those responses. But there was no apparent pattern (data not shown). The one strong association we did observe was with responses to the open-ended question asking them to give an example of their experience with using epidemiology in a risk assessment (“Could you share an example from your experience where you used epidemiologic evidence in risk assessments (or the development of exposure limits)? Did these data make a difference in the risk assessment outcome, whether they confirmed or contradicted the other available data?”). This observation was data-driven, not hypothesized in advance.

We (four independent coders) coded their responses to that question based on positive and negative assessments as follows:

- Positive = described a case of epidemiology trumping other evidence or offered other very positive words about its value
- Neutral = mentioned using epidemiology neutrally, with no praise about it being useful nor any complaints
- Negative = described a case of epidemiology being uninformative, or asserted that it harmed the assessment by introducing misleading inputs

If a respondent mentioned multiple experiences using epidemiology, we coded based on the most positive. Respondents who said they did not have an example to offer, or who reported using epidemiology only in a “hazard identification” or “hypothesis generating” type role are omitted. Note that the details that allowed for this coding were volunteered; we only asked for examples without prompting for an opinion about that experience. As shown in Figure 2, those reporting a positive experience using epidemiology were more likely to be in the “contented” region and were more likely to rate the potential of epidemiology high. Those reporting negative experiences were distributed widely in terms of perceived potential, but consistently negative about current practice.

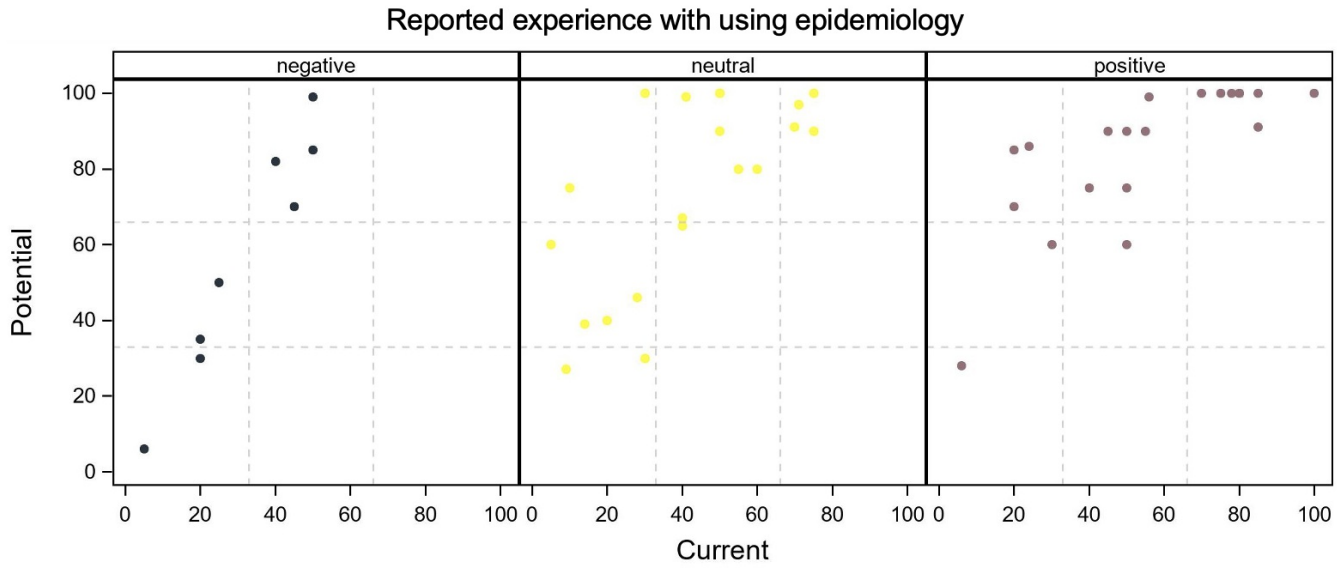

Figure 2. Distribution of rated values of epidemiology in risk assessment, by respondents' reporting of experiences with using epidemiology

These results suggest that global opinions about the conduct of epidemiology are substantially influenced by having a specific experience with a particularly useful or particularly troublesome bit of epidemiology in their own work. This, along with the simple fact that the ratings are so widely distributed, suggests a potential for mutual learning within the risk assessor community, not just between them and epidemiologists. Those who are contented might broaden their understanding of the limitations by hearing details of stories from those with negative experiences, and the pessimists might gain some appreciation of the potential from the positive stories.

Thematic analysis of open ended questions: what epidemiology studies should do

Three open-ended questions asked respondents:

- What do you find helpful in epidemiology papers when you conduct risk assessments?
- What do you find unhelpful?
- If you could ask anything from an epidemiologist to help you conduct a risk assessment or develop exposure limits, what would it be?

Most answers turned out to be interchangeable across these questions, with different respondents choosing to describe a particular study property to be helpful, or its lack to be unhelpful, or it to be something they would ask for. Thus, we analyzed the responses to these three questions together and present the collective results. (It appears that it was useful to ask all three questions, because most respondents identified different study properties across questions.) In addition, some content from other responses was effectively answers to these questions (responses to the question about an example of using epidemiology and the "If not, can you tell us why?" followup to "Have you ever considered using epidemiologic data in risk assessment?"), and so are also included.

We make some observations about the quantity of mentions when they seem informative. Frequent mentions seem to

indicate a problem. However, lack of a mention should not be interpreted as satisfaction. Undoubtedly, if asked to answer questions like “The outcome measure should be clearly defined. Agree/disagree?”, there would have been universal agreement, whereas only about half chose to mention this issue spontaneously. Moreover, it is important to stress that survey respondents cannot be expected to provide comprehensive answers to broad open-ended questions. Failure to mention a topic does not imply a lack of concern, let alone a lack of awareness that it deserves attention. We primarily interpret a respondents’ choice to mention a concern as an indication that they would prioritize it in a discussion of how to improve the epidemiology.

In some quotations, we have corrected obvious typos or changed capitalization, punctuation, or conjugations to improve readability, leaving these unflagged to maximize readability.

An attempt to measure what we really want to know.

Any discussion of the value of epidemiology in risk assessment should start with the observation that though it is messy, difficult, and *ad hoc*, epidemiology is the only way to measure the actual effect of realistic doses of an exposure, and actual outcomes of interest in the species of interest and potentially across identifiable subgroups of that species. As one respondent put it:

“Reported data are directly relevant to real life conditions of the human person or population. No interspecies extrapolation issues, no unrealistic/irrelevant/hypothetical conditions of exposure.”

In addition, there are recognized advantages to epidemiology practice that could theoretically be replicated in lab studies, but rarely are because of how the latter are done. In particular:

> *“The levels that are causing risk in a population that is being exposed to a particular contaminant at very low concentrations for a long time”*
>
> *“To understand the health effects at environmental exposure levels, if any, as animal studies were typically exposed at higher levels.”*

Only a small minority of respondents invoked any version or aspect of these core advantages of epidemiology in their answer about what is helpful. Presumably others thought it was obvious and went without saying. But it is also possible that the small number of mentions represents a failure to recognize just how important these advantages of epidemiology are.

Many respondents who specifically criticized proxy measurements in epidemiology in other categories of responses seemed to not recognize that the alternatives are studies where *everything* is a proxy, including species, doses, dosing schedules, and often outcomes. Others suggested that lab results should trump epidemiology observations and that epidemiology results should be dismissed when they exist without supporting lab results. For example, one respondent commented, “Unfortunately human data often take priority over solid evidence to the contrary in terms of plausibility.”

There were other comments that suggested that epidemiology should clearly trump the proxies. Some of these were cases where there is a perception that no useful proxy model exists or when epidemiology results show that whatever mechanism causing an outcome in rodents is absent in humans. A few (though only a few) suggested the respondent believed that epidemiology is always more informative if it is high quality. It might be interesting to explore how much these contrasting views were attributable to contrasting experiences -- with cases where the epidemiology seems to be faulty versus cases where there is good reason to believe it trumps -- and the extent to which they are based in underlying theory.

These observations relate to the two-dimensional assessment of measurement quality whose axes are sometimes described as “relevance vs. rigor” or “accuracy/unbiasedness vs. precision”. It is possible that a surprising epidemiology result is due to random error or uncontrolled confounding, or to publication bias and model shopping, and should be considered meaningless given a lack of other support (these themes are discussed more below). But it is possible that the surprising result is real given that epidemiology is measuring somewhat different phenomena (the actual effects of actual exposures on *H.sapiens*) and does sometimes discover a causal relationship before the mechanisms are recognized. Interdisciplinary dialog might lead to a more balanced recognition that surprising epidemiology results should be treated (by both epidemiology authors and risk assessors) as somewhere in between the convenient extreme assumptions of error and revelation.

A pessimistic appraisal of the *potential* of epidemiology is tantamount to saying that despite its advantages compared to other assessment methods, there is no way that epidemiology can ever be done well enough to provide useful information. This is a strong claim. It would be interesting to see how many of the “pessimist” respondents would endorse the explicit version of their implicit assertion.

Exposure measurement and characterization

The most common theme in the responses was need for useful exposure measurements and reporting. Almost all respondents mentioned some concern about the exposure variables in epidemiology studies. Basically all those mentions were criticisms of epidemiology conduct, citing something that needs to be done right for the results to be useful, but is often not.

If an epidemiologist were asked to identify the main challenge here, most would probably mention the challenge of finding a variable that is an unbiased and precise estimator of the real exposure of interest. Dissatisfaction with this was indeed mentioned by our respondents (e.g., “Studies in which previous exposure is claimed because of a remembered single or occasional contact in the past only (and even worse when someone else such as a relative has remembered) are mostly not reliable.”). However, the responses more often focused -- rather adamantly -- on simple inadequacies in how the measures were reported and quantified, including:

> *“[Report] concentration, dose, details on substance”*
>
> *“[Report] levels, dose-response information, substance identity, duration of exposure (as a single exposure &*

> *numbers of days/week, number of years, etc.), numbers of subjects affected, temporality of onset of effects compared with exposure."*
>
> *"The crucial point is that actual exposure of study participants to the substance of interest must be verified to the extent possible and efforts should be taken for its quantification. Quality of the study depends very much on the latter."*
>
> *"Adequately characterize exposure (route and all possible sources), if biomarkers of exposure are used coupling those to PBPK [physiologically based pharmacokinetic] models to estimate external dose, and using well established criteria to evaluate causality of effects."*
>
> *"… detailed description of exposure assessment methods, including their validation and if possible estimation of contribution/non-contribution of personal protective equipment in the internal exposure (such info would help in convincing toxicologists)."*

The reporting aspect of these concerns is an easily remedied problem. The epidemiology study authors obviously know (or should know) the answers to those questions, and such reporting is a minimal requirement for doing proper science under any circumstances, let alone when those who should be the target audience specifically need to know it. Perhaps providing honest concrete reporting of exactly what was measured would cause some studies to be ignored by risk assessors as uninformative because those measures fail to provide useful information, but that is as it should be. The failure to clearly report leaves skeptical risk assessors reasonably suspicious that the measures and methods are invalid and the epidemiologists are trying to hide that. It also probably causes some naive risk assessors to blindly accept unsupported conclusions.

The discussion of what measures and methods are valid can really only begin after risk assessors gain a more complete picture of what measures and methods epidemiologists are using (i.e., after the latter report those better). However, some common approaches to exposure assessment are clearly seen as unhelpful. Concerns about some of the exposure assessment practice in epidemiology included:

> *"Generalizations, grouping a category of chemicals with no way to discern exposures to individual chemicals, unreliable exposure characterization methods (such as ever/never use, self-report, etc.), use of spot biomonitoring samples to characterize chronic exposures."*
>
> *"Exposures are often not known with sufficient accuracy to be helpful in the quantification of safety limits."*
>
> *"The crucial point is that actual exposure of study participants to the substance of interest must be verified to the extent possible and efforts should be taken for its quantification. Quality of the study depends very much on the latter. Most suitable (but seldom found) are studies (mostly prospective) that combine examination of adverse health effects with human biomonitoring data."*

Of course, suggestions about what should have been done are substantially improved by an understanding of what *could* realistically have been done given resource constraints and sheer practical impossibility. A better understanding of what could be done better, and how to find an optimal second-best when the ideal is simply not an option, could come from interdisciplinary dialog among epidemiology researchers and risk assessors, making sure to include experts in relevant fields such as exposure science.

How to report exposure levels

Many responses specifically called for stratifications of exposure levels that can inform dose-response assessment. These usually stated or had a subtext that the most useful statistic for risk assessors is an estimate of the no-effect threshold value. One response that discussed this at length argued that epidemiology researchers who want to be helpful should focus on providing their own estimate of this. More often, the responses stressed the need to provide the statistics that allow a reader to assess this, with adequately stratified results. The common epidemiology practice of dichotomizing the world into exposed versus unexposed (or, as one respondent complained about it “Ever/never exposure classification”) is not helpful. Similarly, dividing exposures into a few arbitrary ranges is still inadequate, and is rendered worthless when the ranges are not even reported (“undefined exposures, such as low/med/high, such that results from one study can't be compared to another quantitatively”).

Some responses suggested that strata should not just divide up whatever data happens to exist, but rather should usefully divide the range of greatest interest. One response called for providing more than one alternative stratification to further aid in assessment of no-effect thresholds. There were calls for reporting descriptive statistics and important covariates separately for each stratum (something that the authors can easily do but the reader has no way of determining otherwise), to allow the reader to get some idea about whether there is obvious potential confounding.

There was adamant dissatisfaction with regression models. These models, which epidemiologists use to simplify all the complexity down to a single quantitative estimate (perhaps just because they learned that is just how you always do epidemiology), lose an enormous amount of the information that risk assessors want. They are also heavily assumption-laden, and some of the usual assumptions are flatly wrong, including often precluding the usual core expectation of risk assessors that there will be a no-apparent-effect threshold. Opaque modeling and variable definitions make results especially useless. As respondents put it:

> *“A single regression coefficient...does nothing to help describe what the actual shape of the dose-response is nor does it tell the risk assessor anything about where the POD might be located.”*
>
> *“All the dose-response information in the study is lost in this analysis and these results only provide a Y/N answer to whether there was an effect or not and some rough guess about the dose-response assuming a (likely unrealistic) linear relationship from 0 exposure (control) to the highest exposed. [This is] not very informative for risk assessors, especially if it was known that there were no effects in the lowest exposure groups and effects only began at higher exposures. But this...approach is precisely what many (most) environmental epi studies do with*

> *multivariate linear regression, presenting only a single regression coefficient to characterize the dose-response relationship over the entire exposure range of study participants as a single slope (the coefficient). ”*
>
> *“For example, stating that cholesterol increases by x% for every doubling of serum level of [a particular chemical exposure] is not helpful. Providing the quartile serum levels and the cholesterol measurement for each quartile is what is essential for dose response.”*
>
> *“[It is unhelpful] to report regression coefficients for changes in unusual transformations of outcomes which are uninterpretable to readers without additional info (e.g. beta corresponding to a change in the negative inverse of a test score: what does this mean?)”*

A few respondents offered fairly detailed proposed changes in methods, such as:

> *“[Helpful are] analyses that either model exposure as a categorical variable (e.g. quantiles, as many as possible based on the study size), or using non-linear methods (e.g. restricted cubic splines, GAMs [generalized additive models], etc.). These alternative methods allow the study authors to comment on the shape of the dose-response, suggest a PoD [point of departure] and also allows risk assessors themselves to assess the dose-response when reading and interpreting the study.... Other information from epi studies that are informative for risk assessment includes descriptions of the study sample (distribution of outcome, exposure, covariates) so that risk assessors can determine the relevance of the sample to their own target population for whom they are risk managing exposure.”*

Frustration with these problems of exposure reporting led some respondents to metaphorically throw up their hands and ask for the raw data:

> *“It would be ideal if de-identified raw participant-level data from epi studies were made available to gov’t risk assessors upon request in order to validate and confirm reported findings.”*

It appears that there are some perception here that could be cleared up by active interchange, one that goes in the opposite direction of the emphasis of this paper: Raw epidemiology data is typically not quite so easy to make sense of as these requests imply, and though unadjusted tallies should be reported, they can often be quite misleading without proper social-science methods.

Given the prevalence of stated concerns about exposure, this appears to offer the lowest-hanging fruit for epidemiologists to improve the value and perception of their work among risk assessors. Adequate reporting is a trivial fix, an easy way for epidemiologists to demonstrate they are listening and trying to do useful work. Adequate stratification is similarly easy to implement. Some of the concerns about exposure measurement may be intractable (e.g., a better measure of an exposure is simply not available), but may be worth trying to collaborate to work out, as one response suggested, “how can we leverage current exposure data to better determine the pattern of exposures?”).

Outcomes measurement and reporting

Far fewer respondents mentioned concerns about outcome variables compared to exposure variables. This might mean there are fewer problems, but might reflect respondents not recognizing similar problems. Risk assessors will naturally notice when the exposure measure is inadequately or non-usefully reported or defined, is seemingly poorly measured, or is a bad proxy for the actual exposure. But risk assessors without a background in epidemiology may not even realize how often the same concerns exist for the outcome measure. Toxicology experiments usually have clear bright-line endpoints (albeit ones that are often poor proxies for the human health outcomes of interest), so it may not occur to many who are primarily familiar with such research that epidemiology outcomes can be as slippery as exposures.

Most respondents who did mention concern about outcome measurements were concerned that the measures were often "risk factor proxies" rather than "real health endpoints", as one of them phrased it. The perfect solution to this -- e.g., knowing how many people in the study cohort will be diagnosed, over the next few decades, with a disease of concern -- is simply not practical to hope for. Some interdisciplinary dialogue is needed about what is possible and what amongst that is most useful. Other comments on this topic included calling for:

> *"Quantitative exposure and response data, careful and specific definitions of outcomes and exposure conditions."*
>
> *"Transparency in... assessments done"*
>
> *'Interpretation or explanation of the importance or relevance of the health outcome.... Is [it] meaningful?"*
>
> *"Clear indicators of the issues, direct link to organ system or outcomes.… Does the endpoint in question result in a lowering of life expectancy?"*

Biomarkers

A large portion of the comments about outcomes included a call for biomarkers, as did many of the comments about exposures. For example:

> *"external dose information paired with internal dose information.... Associations that relate back to quantifiable external exposure levels."*
>
> *"If there is no biomonitoring data, then proxy measures of exposure...must be relied upon, which do not indicate true internalized exposures and cannot be relied upon in a quantitative risk assessment."*
>
> *"Adequately characterize exposure (route and all possible sources), if biomarkers of exposure are used coupling those to PBPK models to estimate external dose, and using well established criteria to evaluate causality of effects."*

There was an absence of specifics about *what* biomarkers might be useful, not even examples. This is not surprising given the broad questions and lack of further probing of themes. Still, we got the impression that many respondents perhaps did not even know what might be realistic to hope for considering costs and logistics, and so this is a potentially fruitful area for more mutually educational dialog. Only a few of those who mentioned biomarkers gave an indication that they recognized the epidemiologic challenges of relating the exposure biomarkers that might be measured to external (environmental) exposure levels:

> *"difficult to convert from blood concentration of [chemical] to dose relationship"*
>
> *"unreliable exposure characterization methods [include] use of spot biomonitoring samples to characterize chronic exposures."*

There are challenges in doing and interpreting many biomarker measures that are not necessarily well-known among risk assessors, who may think any internal biological measure is always good to use. More conversation between risk assessors and epidemiologists might help both groups better understand what is genuinely most helpful in particular circumstances.

Study types

Calls for biomarker measurement were frequently coupled with calls to do more prospective cohort studies, which are needed in order to do biomonitoring (unless there is optimal biobanking, which was also called for, though this tends to require unrealistic precognition). Comments included:

> *"If the question is asked to an epidemiologist prior to the conduct of a study, then the request would be to see more comprehensive prospective cohort studies leveraging data from a large group of people. Ideally, these studies would contain multiple time-points, long periods of follow-up, with corresponding biomonitoring data. These studies would overcome the disadvantages of cross-sectional and ecological study designs, which are unable to assess temporality."*

Prospective cohorts were always the preferred study type when an opinion was offered, but the reasons given often reflected misunderstanding. Resolving such misunderstanding might not be accomplished via interdisciplinary dialog with just any epidemiologists, unfortunately, because many epidemiologists have erroneous beliefs about study types. The epidemiology experts would have to weigh in. In particular, some respondents invoked the unfortunately common erroneous claim that some study types actually show causation while others do not (but only "suggest association but never establish causation"). For example, "ecological and cross-sectional study designs are unhelpful as they are often only exploratory in nature and were not designed to assess a direct cause-effect relationship." The reality is that causation is never observable, and can only be inferred with varying degrees of confidence. It is certainly true that some studies yield greater confidence that an association is causal, but that is affected more by the quality of the data and analysis than

the study type. There are a few studies in epidemiology (predictive modeling) that are explicitly not trying to assess causation, but basically every study in the risk assessment realm is trying to assess causation, whether or not they do a good job of it.

Perhaps the most obviously mistaken belief is the notion that retrospective studies cannot gather information on the timing of past events. This is contradicted by the simple fact that retrospective studies routinely measure when someone was born without difficulty. Causal inference is not dependent on whether the study starts before the outcomes occur. More subtle is the lack of recognition that some endpoints and exposures (especially biomarkers of exposures and outcomes) can practically only be measured in cross-sectional designs that allow for intense contact with participants.

It was suggested that prospective cohorts allow for assessing multiple outcomes (as is often done in animal studies). This Is valuable, though other study designs can also do this.

This limited survey does not allow detailed assessment of what respondents understand about epidemiology study designs, but is enough to suggest topics for future conversation. There may be a widespread lack of understanding that while repeated cross-sectional designs are not cohorts like are ideal for cancer epidemiology, they are a better approach for some questions. Most prospective study designs fail to fully predict unfolding knowledge, so retrospective information is still gathered over the many decades they take to come to fruition. Prospective cohorts are often hopelessly inefficient (as with combinations of common exposures and rare outcomes) and can only happen under limited circumstances.

It could be useful to ask risk assessors to tell epidemiologists what details of retrospective cross-sectional studies would be most useful, after prompting them with the reasons why cancer-epidemiology-style cohorts are seldom a practical method. This seems like a particularly ripe area for bilateral education. One response suggested a launching point for this:

> *“Think more mechanistically. An understanding of underlying toxicology and using it to create context would help...retrospective analyses considerably.”*

Study populations

There were relatively few mentions of study population issues, and most of those focused on needing an adequate sample size. This might indicate general satisfaction, but it might be that some respondents overlook potential concerns. Someone with a background in experimental sciences -- most familiar with randomizing exposures to produce an unbiased effect measure for one selected population -- may be unaware of the challenges in epidemiology to avoid bias from sample recruitment for a supposedly representative population.

A few respondents did comment beyond sample size. There was a call for better reporting of “descriptions of the study sample (distribution of outcome, exposure, covariates) so that risk assessors can determine the relevance of the sample to their own target population”. One respondent asked epidemiologists to assess, “What are the key issues that impact on the extrapolation of the results of epidemiological studies to the whole population?”

There were specific requests to make sure to include target identified vulnerable populations, and to report results for the other target demographics (age, sex, race, etc.) that risk assessors are asked to address, and to try to sort out the most sensitive populations. This is something that epidemiology is uniquely able to do, but apparently often disappoints. The comments about sample size did not include specifics about targets (e.g., “adequate” to do what?) or which subpopulations needed to have adequate sample size (we can infer that this would include having subjects close to the threshold exposure level). Operationalizing these requests is more complicated that is immediately apparent, and seems a fruitful area for dialog.

Confounding

Half the respondents expressed concerns about the handling of confounding in epidemiology studies (though few used the word “confounding” as epidemiologists almost always would, which is perhaps subtle evidence of a lack of communication). The most common concrete concerns were about exposures to mixtures, subjects’ chemical exposures other than the target chemical of interest (“Multiple exposures to various chemicals makes determination of causation difficult.”). There was similar concern about whether the blame was rightly placed on the exposure itself rather than an effect of it (“a detoxification product or an adaptive response”) though it is not clear that this is a widespread concern. A few comments had a theme that epidemiology has confounding, so that makes it always inferior information (aka, the “aha, I found potential confounding so we must dismiss this study” nihilism), though we were surprised, based on our previous experience, how seldom such comments appeared. There was some explicit optimism about handling confounding (e.g., “Clear improvements have been achieved over the course of the years.”), but not unexpectedly, no responses delved into any specifics on how to accomplish this.

Few of the mentions of confounding issues came with any concrete suggestions, perhaps reflecting a recognition that all randomized studies are alike, whereas every study with systematic confounding is confounded in its own way (and thus the epidemiology researchers are the ones who need to figure out how to deal with it). Several of the comments simply said “adjust for” it, which is rather too optimistic about epidemiology methods. Unfortunately, this mistaken impression about what epidemiology can do reliably is not likely to be remedied by interdisciplinary dialog, since most epidemiologists share this pollyannaish view of controlling for confounding. Perhaps the most important points regarding confounding can be found in the theme about asking for more honest reporting of study weaknesses (below).

Establishing causation

There were various vague calls for clearly demonstrating that observed associations were causal, and a few outright rejections of epidemiology because it can “never establish causation”, an even stronger version of the aforementioned erroneous perceptions about study types. A bit of interdisciplinary dialog might not do much to address the underlying fundamental issue about the nature of causal inference in scientific inquiry because the average epidemiology researcher also lacks an understanding of this concept (see, e.g., appeals to fallacious “causal criteria”). Better-than-average expertise (and a deeper dive than the present project allows) would be required to optimize the dialog. However, there are many things that all epidemiologists can do to help their readers (and themselves!) better assess how much causation

should be inferred, including most of what appears in the following two themes. Dialogue can probably be fruitful in defining causal questions, because scientists with different disciplinary backgrounds tend to anchor their inquiry in different causal questions (e.g., “does exposure cause disease in the experimental system” vs. “does exposure cause an epidemic among humans”).

Epidemiologists, please just tell us what the heck you actually did!

The section heading is a paraphrase of a sentiment that many of the respondents endorsed. This was a recurring theme that appeared, implicitly or explicitly, in a large portion of the comments. As already noted, it appeared in the comments about exposures and outcomes. It was also embedded in comments about covariates, confounding, and population selection and recruitment. In sixteen cases it appeared as a freestanding general admonition. For most of the topical themes, including exposure reporting, outcomes reporting, and confounding, there were more critical comments about the failure to report what was done, fully and properly, than there were assertions that it was done badly. Opacity was a bigger perceived problem than any method that was sufficiently explained to be judged as poor. One respondent reacted to the question about unhelpful epidemiology by saying that it is not so much that epidemiology researchers’ choices were unhelpful as “missing analysis/information or study limitations in the papers...decrease the usefulness”.

Many of these comments reflect the failure of epidemiology authors, far from unique to this subject matter, to simply communicate something they know and should state. E.g., what exactly was the exposure/outcome that was measured (and how, exactly, it was measured, including providing the actual questions used for survey data). Others require some additional calculations which clearly should exist, and that the authors could easily do but the reader cannot. E.g., statistically characterizing the strata or any subpopulations that are reported, rather than only giving those statistics for the whole population. In keeping with a previous observation, a few respondents seemed to despair of steps like these ever being done properly and said just give us your raw data and we will do it ourselves.

Comments included:

> *“provide a very transparent reasoning on its study design, data collection and interpretation of the findings”*
>
> *“[Avoid a] lack of detail [such as] results presented only as figures.” (Every reader who has had to use a ruler to estimate quantities to do calculations that authors failed to report will sympathize.)*
>
> *“The availability of online supplemental material makes it more feasible to present fuller information of methods and results - the more information we have the better we can make our risk assessment. In some cases, exploring and reporting BOTH joint and independent effects (e.g., for highly correlated pollutants) is very helpful.”*

These are complaints that readers of almost any topic in epidemiology could level at the authors, but are seldom voiced with much vehemence. Universal improvements in education and standards in the field are warranted. Perhaps those could be motivated by readers like risk assessors who cannot just give up and accept vague general conclusions when they cannot figure out exactly what a result means.

Several respondents emphasized the need for adequate reporting because epidemiology is almost always done*ad hoc*, with study methods inconsistent with one another:

> *"Since there are often no definitive test guideline requirements for epidemiology studies, it is unhelpful when there is a lack of transparency in data analysis or methodology, which contributes to difficulties in reproducing results between studies."*
>
> *"While you may not agree with the principles presented in existing risk assessment guidelines for specific endpoints, at least you know that standardized criteria have been established to guide an assessment. Study quality and usefulness can be judged against how well the study authors have conducted their assessment against those criteria. With epi studies, it's a bit of a c--- [sic] shoot."*

These are very valid points. Epidemiology studies use bespoke methods, and if these are not fully reported, the reader simply does not know what was done. It is a general problem, a failure of basic scientific practice, that most epidemiology reporting implicitly tells readers to simply accept their results without explaining what was done, and they usually get away with it.

Only seven respondents commented specifically on failures to explain statistical methods or report the internal assumptions in models (e.g., "Strange dose-response curves determined by non-transparent statistical data treatment"). Since these failures are nearly universal in epidemiology reporting, the small number of mentions seems to suggest most respondents do not dig far enough into the epidemiology reports to even realize what is missing. Perhaps this is because the more common complaint is that most complicated statistical methods render the results useless and should be avoided entirely, and thus there is limited motivation to worry about how they were done.

There were few comments about the incommensurability of results across epidemiology studies. Such comments tended to be about a specific point (e.g., "undefined exposures, such as low/med/high, such that results from one study can't be compared to another quantitatively" and "lack of transparency in data analysis or methodology...contributes to difficulties in reproducing results between studies"). Problems like those can be fixed *in situ*, but the bigger picture problem requires a bigger change in epidemiologists' behavior: Studies that epidemiology authors imply replicate a previous assessment almost invariably fail to do so. They ostensibly look at the same exposure and outcome, when those are described in simple casual language, but they invariably use different measures, functional forms, and statistics. Sometimes this is unavoidable because genuinely replicating data does not exist. But even to the extent it is possible to make a later study replicate previous work -- such as by using similar models, covariates, variable definitions, etc. -- it is seldom done, and every study takes a new *ad hoc* approach. Perhaps many risk assessors do not even recognize that such real replication is not being done and would call for it if they realized it (and a few even asked for meta-analyses, which is not possible to do meaningfully unless the results can be made commensurate).

Calls for balanced and careful assessments of study strengths and weaknesses

Taking the previous theme one step further, about a third of responses explicitly called for epidemiology authors “to provide an honest critique of the strengths and limitations of the study”, or some equivalent. The most common context was dealing with potential confounding, perhaps reflecting how this is largely opaque to the reader unless there is a concerted effort to honestly present it.

> *“[Need] discussion of risk of bias, confounding, covariants, etc. In the absence of clear disclosure of those issues, plus open discussion of the strengths AND limitations of each study, epi data for use in risk assessment is unreliable.”*
>
> *“What external factors or parameters that were not included in the study may have influenced the end result?”*

Other comments included:

> *“Provide a better description of uncertainty of exposure assessment”*
>
> *“Detailed explanation of the uncertainties in the study data; as a toxicologist, I would appreciate an interpretation (from an epidemiologist) of the results.”*
>
> *“[Provide an assessment of] confidence in the results...so I can decide if I am able to use them.”*

Indeed, it should not be wholly up to the reader to try to assess these. The study authors ought to be able to contribute useful assessments and they are much more able to do some of the analyses that might inform the answers, such as assessing how sensitive the result is to the exact effects of a confounder or other recognized source of uncertainty.

Reporting uncertainty is, of course, inherently part of proper scientific practice, but it is vanishingly rare in epidemiology. There is almost never an attempt to quantify uncertainty beyond random sampling error (which leads naive readers to believe that is all the uncertainty, though those in our survey population are unlikely to make that mistake). “Study limitations” paragraphs are generally useless, treated more as a confessional (list the study’s sins and consider them absolved and thus ignorable) or worse (mention trivial concerns to try to distract from fundamental problems). This is never adequate, though few readers push back. It was striking to see the frequency of explicit push-back in the survey responses, though this should be expected from readers who need actual scientific quantification, not epidemiology papers that feel like newspaper articles.

At least do no harm

The most negative reports of experiences with epidemiology being used in risk assessment referred to cases where the respondent believed epidemiology conclusions actively interfered with good risk assessment. Of course, it is possible that some observers might conclude that the epidemiology really corrected an erroneous assessment that was based on other science; we do not have enough information to judge that. But we can conclude there are major disconnects, and that letting one science simply trump the other as a result of personal bias or successful rhetoric -- rather than pursuing a

synthesis of all we know -- is not a good outcome.

Other comments about epidemiology causing harm appeared across the open-ended responses. There was concern about poor quality epidemiology interfering with rather than improving risk assessment:

> *"Conduct a high-quality study or don't bother!"*

A few respondents commented on particular behaviors of epidemiologists, such as "splitting the population in multiple ways to try to find some positive association" or other tactics for choosing functional forms to try to manufacture a preferred estimate. A related problem is "selective reporting: only reporting selective 'statistically significant' or interesting results and not reporting other results", habitually failing to "present results of all analyses, negative or positive". This hidden model shopping behavior (aka, publication bias *in situ*, p-hacking) is a common problem in the field, despite being frankly unethical conduct. It biases both the results and any estimates of uncertainty (by creating a systematic bias that cannot be accounted for in either formal or intuitive methods of estimating uncertainty). This problem is an aspect of a previous theme also, the need for complete reporting of methods and meaningful sensitivity analysis, specifically including reporting the results of alternative models that were tried, along with a justification for continuing to model shop and for choosing the particular model to highlight. It seems likely that most respondents are unaware how common and problematic model shopping behavior is, and that they would become (justifiably) more doubtful about the current quality of epidemiology if they realized.

Some respondents categorically objected to the *ad hoc* data-driven analyses that are common in epidemiology studies. For example:

> *"...need a clearly defined study plan... [without any] after-study mathematical unplanned treatment of data."*

Adhering to this frequently-stated prescription for scientific research is certainly one way to avoid the common problems caused by hidden model shopping behavior. But there are compelling arguments that epidemiology is more useful if some of the analysis is based on data distributions or other observations that were unknown during study design. Unfortunately, epidemiologists frequently take this too far, and they fail to report results from alternative and seemingly equally-valid (or more valid) modeling choices. Perhaps some interdisciplinary dialog could lead to a mutual recognition of how epidemiologists can introduce the right level of data-drivenness into their results. At least it would educate epidemiologists about how their results are often (quite reasonably) ignored because of suspicions about aggressive model shopping.

Several respondents commented that epidemiology authors frequently opine worldly conclusions or qualitative policy recommendations without doing any relevant analysis beyond their single study (e.g., "Conclusions that are not traceable from the numbers presented"). These included specific criticisms of over-concluding based on mere statistical significance, particularly in contexts where there is a lot of apparent confounding. There was blunt advice that the results are potentially useful, but epidemiology authors' conclusions are not. Only a few respondents went quite so far, which

might be because this is such a harsh condemnation of epidemiologists. But we might instead infer an even harsher conclusion, knowing our own behavior: some and perhaps most experienced risk assessors simply ignore the "conclusions" section of papers, expecting them to offer nothing of value, and so do not really care about their quality. A few responses advised epidemiologists on what form their conclusions should take (e.g., "Does the end point in question result in a lowering of life expectancy, total years."), so at least some risk assessors seem to think encouraging epidemiology authors to craft useful conclusions is a possibility.

A few respondents railed against cases where the epidemiology researchers do not leave their unanalyzed conclusions as throw-away sentences in the paper, but broadcast them:

> *"Study authors are often reckless in suggesting causation and media even more so. This is a serious headache for regulators."*
>
> *"Can we be realistic about risk assessment, and stop playing fear factor with all of these situations?"*

One response noted specifically that sometimes other evidence strongly suggests that a particular exposure is too low to plausibly have an effect, but "many low exposure studies appear to be agnostic about observing a potential effect...and when they do observe effects they are frequently considered to be reasonably causal associations even when...the literature suggests it is very implausible or unlikely." This relates to the topic of the next subsection.

Integration of other scientific knowledge, mechanisms and collaboration

Another practical and operationalizable theme involved asking the epidemiologists to better attend to other scientific knowledge about the topic.

Some of this was simply a matter of making sure the epidemiologists were even aware of the other science and had some notion of how it related to their work:

> *"collaborate with experienced exposure scientists and toxicologists"*
>
> *"have discussions during the study design phase of their work"*

Drilling down, some respondents got more specific and argued -- plausibly -- that without the mechanistic understanding that comes from reviewing the previous science, it is unlikely that epidemiologists will design their studies optimally:

> *"Think more mechanistically. An understanding of underlying toxicology and using it to create context would help...retrospective analyses considerably."*
>
> *"Epi work should be considered along with biological, toxicological data. Because of the multiple confounders involved in trying to evaluate and obtain epi data, the results should be examined carefully. Epi data should not*

> *stand alone but be considered as an adjunct to properly performed biological and toxicological information. Epi data must have a biological foundation to be considered useful and all confounders need to be considered and revealed."*

One respondent identified their reason for not considering epidemiology when doing risk assessment as the absence of being able to find "realistic links between animal studies and what may have been observed in human populations, and how those data might be used in risk assessment." Some of these sentiments are a bit too dismissive of the possibility that epidemiology can discover an important causal relationship for which there is no existing mechanistic story. But epidemiologists should realize that making such a genuine discovery is so rare (probably less than once-in-a-career) that it does not give them an excuse for ignoring the other knowledge, especially in the context of ongoing risk-assessment which tends to generate a lot of other analysis.

Related to the theme of providing a clear assessment of study limitations, some respondents suggested that epidemiologists need to offer some explanation of how their results integrate with previous knowledge. In particular, they should acknowledge when the claimed study results are contrary to what is suggested by other evidence, and assess and explain why that is.

> *"Are they familiar with similar results from other epi studies? How confident are [the authors with] the results?"*
>
> *"[Report] the comparison between the limit value derived from toxicological data and human data and in case of high differences [assess] the meaning of them."*

One of the themes that offers a simple solution is wanting collaboration between those who generate epidemiologic results and those who use them, at the design phase of the epidemiologic research. It seems like this ought to go without saying, but is clearly missing from epidemiology practice according to our respondents. Epidemiologists not even bothering to talk to the people who know the biology is a terrible misapplication of resources.

> *"Team work with exposure assessors, toxicologists."*
>
> *"having discussions during the study design phase of their work"*
>
> *"Epi data should not stand alone but be considered as an adjunct to properly performed biological and toxicological information. Epi data must have a biological foundation to be considered useful and all confounders need to be considered and revealed."*

Relating that sentiment to the previous theme of doing no harm,

> *"Epi studies are a good tool, but should only be interpreted as part of the puzzle, not the word of truth and facts."*

Emergent overarching themes

An emergent theme from many comments is that epidemiologists have more confidence in their field and their findings than is warranted by their tools, methods, and data, and this makes their work far less useful. Complaints relating to this include epidemiology researchers wanting readers to just blindly accept that they usefully measured the exposure or outcome they claim to be reporting; expecting the reader just accept as valid whatever model they happened to create; ignoring the context of other research on the topic; failing to identify and discuss serious concerns about the study; and not attempting to present results in terms that the most important target audience can make use of. The current state of epidemiology is definitely not a field that justifies arrogance. Perhaps hearing the input from those who would like to make use of their work, but have a difficult time doing so, would help temper this problem.

We originally interpreted another emergent theme as an unrealistic request to make epidemiology more turnkey. Recipes for translating lab data into, say, estimates of no effect levels are precise and easy to implement (though, of course, are frequently inaccurate due to everything being a proxy). Demanding that inherently more complicated epidemiology analyses provide the same simplicity is an unreasonable ask. However, on further reflection, we came to view most of these comments as a call for discussion in a context of not quite knowing what to even discuss, rather than an unreasonable expectation. One respondent stated:

> *"...animal studies have data for no observable adverse effect level (NOAEL), lowest observable adverse effect level (LOAEL), and different exposure levels to conduct benchmark dose analysis to help derive an exposure limit. For epi papers, results are presented as relative risk, odds ratio etc, which are not equivalent to NOAEL/LOAEL in the animal studies for comparison. Traditional risk assessors who are only familiar with animal studies (NOAEL, LOAEL) do not understand how to incorporate epi data quantitatively in their risk assessments. Epi studies ended up used [merely] as supporting studies in developing exposure limits. [Could epidemiologists could tell us how] to quantitatively develop exposure limits using epi data? How can epi studies be used for benchmark dose analysis using the US EPA's benchmark dose software (BMDS) to help develop an exposure limit."*

Risk assessors are not asking that all results fit into a simple algorithm, but are saying they would like to use epidemiology but need the epidemiologists to help figure out what to do with it. Moreover, there is a subtext that they do not quite even know what help they can ask for. Interdisciplinary dialog is an apparent solution to this, both in terms of getting epidemiology researchers to deliver a product that is more useful as already discussed, and in terms of helping risk assessors figure out what to ask for and how to use it. Presumably epidemiology researchers have some vision, at least a vague notion, of how their research should contribute to practical decisions. Perhaps communicating that would help make it happen. And if such communication leaves risk assessors responding that it still does not make the epidemiology useful, that should be a lesson to the epidemiologists about what they are doing badly.

Conclusions

It is apparent from our survey results that risk assessors (at least those at the level of sophistication in our sample population) can identify various ways in which epidemiology can and should be done better to contribute to their work. Many of the fixes they proposed are very doable for almost all relevant epidemiology research, and can be done at the data analysis and reporting stage and thus could be implemented immediately. Indeed, such fixes could be done retrospectively, with researchers re-analyzing and reporting data that they previously presented in a way that rendered their effort largely a waste of resources.

Other improvements would require changes in how epidemiology research was designed from the start, and are not always possible, but would be useful to consider. These straightforward requests and recommendations are an immediate source of value from this exercise in communication. Epidemiologists may not be able to make much practical use of top-down aspirational treatises on risk assessment, but these specific messages are immediately operationalizable advice.

Beyond the low-hanging fruit, it is apparent that there would be benefits from further communication and collaboration. Abstract wish-lists about methods tend to be dismissed as soon as the audience finds themselves saying “yeah, right; good luck with that” about some entry, and the probability of that happening approaches 100% as the list increases in length. But conversations can produce a shared understanding of what is possible and reasonable, and inspire efforts to bring about what can be done.

Epidemiology studies are expensive and risk management policies are more expensive. It is an unconscionable waste of society’s resources to not make the effort to improve the former, to keep them from being useless, and to give them greater potential to improve the latter. Presumably most epidemiology researchers want to contribute to improving policies, believe what they are doing does so, and would prefer to learn when and how this is not the case. One way towards this ideal is embedding the epidemiology in the risk assessment process, creating a closer relationship with risk assessors and with attentiveness from epidemiologists to what risk assessors welcome and endorse.

Statements and Declarations

Acknowledgments

The authors are grateful to all those who completed the questionnaires and helped distribute them. Matt Gribble (University of California at San Francisco) and Mark Johnson (U.S. Army Public Health Center) contributed to early drafts of the manuscript.

Conflicts of Interest

The authors declare no financial conflicts of interest. IB and CVP disclose activism on behalf of better bias analyses and other improvements in the conduct of epidemiology, belief in existence of thresholds in biology, and sincere desire to make risk assessment more quantitative.

References

- Burns CJ et al. (2019). A matrix for bridging the epidemiology and risk assessment gap. Global Epidemiology, 1. https://www.sciencedirect.com/science/article/pii/S2590113319300057
- Deglin SE et al. (2021). Environmental epidemiology and risk assessment: Exploring a path to increased confidence in public health decision-making. Global Epidemiology, 3. https://pubmed.ncbi.nlm.nih.gov/37635726/
- Federal Focus Inc. (1995). The London Principles. Current version available at: http://www.fedfocus.org/science/london.html
- Goodman JE et al. (2020). “Good Epidemiology Practice” Guidelines for Pesticide Exposure Assessment. International Journal of Environmental Research and Public Health, 17(5114). doi:10.3390/ijerph17145114.
- LaKind JS et al. (2020). Bridging the epidemiology risk assessment gap: An NO2 case study of the Matrix. Global Epidemiology, 2. https://www.sciencedirect.com/science/article/pii/S2590113320300018
- LaKind JS et al. (2023). Epidemiology for risk assessment: The US Environmental Protection Agency quality considerations and the Matrix. Hygiene and Environmental Health Advances, 6. https://www.sciencedirect.com/science/article/pii/S2773049223000156
- Samet JM, Schnatter R, Gibb H (1998). Invited Commentary: Epidemiology and Risk Assessment. American Journal of Epidemiology, 148(10).
- The Spice Girls (1996). Wannabe. https://www.youtube.com/watch?v=Wzxo8sUHYvQ.